Thomas G. Robertazzi

Introduction to Computer Networking

Springer

Introduction to Computer Networking

Thomas G. Robertazzi

Introduction to Computer Networking

Springer

Thomas G. Robertazzi
Department of Electrical and Computer
Engineering
Stony Brook University
Stony Brook, NY, USA

ISBN 978-3-319-53102-1 ISBN 978-3-319-53103-8 (eBook)
DOI 10.1007/978-3-319-53103-8

Library of Congress Control Number: 2017931075

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Rachel and Deanna,

for Making Lives for Themselves

Preface

Networking is a fascinating field. Networking involves communication theory, algorithms, technology and diverse environments and situations in an enticing mix. This introductory book on networking technology is meant to convey some of the excitement and variety that may be found in this area. The qualitative coverage of this book spans the smallest possible networks that are implemented on integrated circuit chips to the largest networks conceived that cover the solar system. There is much in between these two extremes.

This book is intended to provide a technology overview in undergraduate and graduate networking courses where it may be used in conjunction with more algorithm-mathematical-oriented texts. It will also be of interest to the individual engineer, computer scientist or information technology professional. I would note that this book grew out of my 2011 Springer brief book, *Basics of Computer Networking*.

I have learned a great deal about networking by teaching networking courses at Stony Brook. I would like to thank a number of people for suggestions on coverage including Tricia Chigan, Victor Frost, Biswanath Mukherjee, Shivendra Panwar and Ruhai Wang. I appreciate Dantong Yu for making me aware of technological trends over the years. I thank Emre Salman and Wendy Tang for proofreading one section of the manuscript. My work life has been made easier by the office staff here of Cathryn Mooney, Susan Nastro, Tim Higgins and Rachel Ingrassia. I have benefitted from the IT assistance of John Joseph and Tony Olivo. I am grateful for the support of my editor at Springer, Mary James, and the project staff of, Murugesan Tamilselvan and Essoudasse Catherine. I wish also to thank Yang Liu and Li Shi for assistance with manuscript preparation.

Finally I dedicate this book to my two daughters, Rachel and Deanna.

Stony Brook, NY, USA — Thomas G. Robertazzi

Contents

Chapter 1
Introduction to Networks

1.1 Introduction

There is something about technology that allows people and their computers to communicate with each other that makes networking a fascinating field, both technically and intellectually.

What is a network? It is a collection of computers (nodes) and transmission channels (links) that allow people to communicate over distances, large and small. A Bluetooth personal area network may simply connect your home PC with its peripherals. An undersea fiber optic cable may traverse an ocean. The Internet and telephone networks span the globe. Networks range in size from networks on chips to deep space networks.

The Internet has been developed over the past 45 years or so. The 1980s and 1990s saw the birth and growth of local area networks and SONET fiber networks. The 1990s and the early years of the new century have seen the development and expansion of WDM fiber multiplexing. New wireless standards continue to appear. Cloud computing and data centers are increasingly becoming a foundation of today's networking/computing world. Networking is also becoming more software oriented.

The book's purpose is to give a concise overview of some major topics in networking. We now start with an introduction to the applied aspects of networking.

1.2 Achieving Connectivity

A variety of transmission methods, both wired and wireless, are available today to provide connectivity between computers, networks, and people. Wired transmission media include coaxial cable, twisted pair wiring, and fiber optics. Wireless

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_1

technology includes microwave line of sight, satellites, cellular systems, ad hoc networks, and wireless sensor networks. We now review these media and technologies.

1.2.1 Coaxial Cable

This is the thick cable you may have in your house to connect your cable TV set-up box to the outside wiring plant. This type of cable has been around for many years and is a mature technology (Wikipedia). The coaxial cable was invented by Oliver Heaviside in 1880 and operates as a transmission line. The first transatlantic coaxial cable was deployed in 1956. While still popular for cable TV systems today, it was also a popular choice for wiring local area networks in the 1980s. It was used in the wiring of the original 10 Mbps Ethernet.

A coaxial cable has four parts: a copper inner core, surrounded by insulating material, surrounded by a metallic outer conductor, finally surrounded by a plastic outer cover. Essentially in a coaxial cable, there are two wires (copper inner core and outer conductor) with one geometrically inside the other. This configuration reduces interference to/from the coaxial cable with respect to other nearby wires.

The bandwidth of a coaxial cable is on the order of 1 GHz. How many bits per second can it carry? Modulation is used to match a digital stream to the spectrum carrying ability of the cable. Depending on the efficiency of the modulation scheme used, 1 bps requires anywhere from 1/14 to 4 Hz. For short distances, a coaxial cable may use 8 bits/Hz or carry 8 Gbps.

There are also different types of coaxial cable. One with a 50 Ω termination is used for digital transmissions. One with a 75 Ω termination is used for analog transmissions or cable TV systems. The most frequently used coaxial cable for home television systems is RG-6.

A word is in order on cable TV systems. Such networks are locally wired as tree networks with the root node called the head end. At the head end, programming is brought in by fiber or satellite. From the head end cables (and possibly fiber) radiate out to homes. Amplifiers may be placed in this network when distances are large.

For many years, cable TV companies were interested in providing two way service. While early limited trials were generally not successful (except for Video on Demand), in recent years cable TV has winners in broadband access to the Internet and in carrying telephone traffic.

1.2.2 Twisted Pair Wiring

Coaxial cable is generally no longer used for wiring local area networks. One type of replacement wiring has been twisted pair. Twisted pair wiring typically had been previously used to wire phones to the telephone network. A twisted pair consists

of two wires twisted together over their length. The twisted geometry reduces electromagnetic leakage (i.e., cross talk) with nearby wires. Twisted pairs can run several kilometers without the need for amplifiers. The quality of a twisted pair (carrying capacity) depends on the number of twists per inch.

About 1990, it became possible to send 10 Mbps (for Ethernet) over unshielded twisted pair (UTP). Higher speeds are also possible if the cable and connector parameters are carefully implemented.

One type of unshielded twisted pair is category 3 UTP. It consists of four pairs of twisted pair surrounded by a sheath. It has a bandwidth of 16 MHz. Many offices used to be wired with category 3 wiring.

Category 5 UTP has more twists per inch. Thus, it has a higher bandwidth (100 MHz). Newer standards include category 6 versions (250 MHz or more) and category 7 versions (600 MHz or more). Category 8 at 1600–2000 MHz for 40 Gbps Ethernet is under development (Wikipedia).

The fact that twisted pair is lighter and thinner than coaxial cable has speeded its widespread acceptance.

1.2.3 Fiber Optics

Fiber optic cable consists of a silicon glass core that conducts light, rather than electricity as in coaxial cables and twisted pair wiring. The core is surrounded by cladding and then a plastic jacket.

Fiber optic cables have the highest data carrying capacity of any wired medium. A typical fiber has a capacity of 50 Tbps (terabits per second or 50×10^{12} bits per second). In fact, this data rate for years has been much higher than the speed at which standard electronics could load the fiber. This mismatch between fiber speed and nodal electronics speed has been called the "electronic bottleneck." Decades ago the situation was reversed, links were slow and nodes were relatively fast. This paradigm shift has led to a redesign of protocols.

There are two major types of fiber: multi-mode and single mode. Pulse shapes are more accurately preserved in single mode fiber, lending to a higher potential data rate. However, the cost of multi-mode and single mode fiber is comparable. The real difference in pricing is in the opto-electronics needed at each end of the fiber. One of the reasons multi-mode fibers have a lower performance is dispersion. Under dispersion, square digital pulses tend to spread out in time, thus lowering the potential data rate. Special pulse shapes (such as hyperbolic cosines) called solitons, that dispersion is minimized for, have been the subject of research.

Mechanical fiber connectors to connect two fibers can lose 10% of the light that the fiber carries. Fusing two ends of the fiber results in a smaller attenuation.

Fiber optic cables today span continents and are laid across the bottom of oceans between continents. Repeater distances are 70–150 km (Wikipedia). Fiber optics is also used by organizations to internally carry telephone, data, and video traffic.

Wavelength division multiplexing (WDM) systems multiplex multiple optical signals, each at a different optical frequency (or color), on a single fiber to boost

the data per time a fiber carries. For instance, a fiber with 100 channels, each of 10 Gbps, carries a total of 1000 Gbps or 1 Tbps. In 2013 researches implemented a system carrying 400 Gbps over a single channel (Wikipedia).

1.2.4 Microwave Line of Sight

Microwave radio energy travels largely in straight lines. Thus, some network operators construct networks of tall towers kilometers apart and place microwave antennas at different heights on each tower. While the advantage is that there is no need to dig trenches for cables, the expense of tower construction and maintenance must be taken into account. It should be noted that transmissions in the microwave frequencies are also used in other applications such as cellular phones and space communications (including satellite communications).

1.2.5 Satellites

There are about 3600 satellites of all types in orbit, of which about 1000 are operational (Wikipedia). Satellites are used for such functions as communications, navigation, weather, earth sensing, and research. Arthur C. Clarke, the science fiction writer, made popular the concept of using satellites as communication relays in the 1940s. Satellites are now extensively used for communication purposes. They fill certain technological niches very well: providing connectivity to mobile users, for large area broadcasts and for communications to areas with poor infrastructure. Two major communication satellite architectures are geostationary satellites and low earth orbit satellites (LEOS). Both are now discussed.

1.2.5.1 Geostationary Satellites

You may recall from a physics course that a satellite in a low orbit (hundreds of kilometers) around the equator seems to move against the sky. As its orbital altitude increases, its apparent movement slows. At a certain altitude of approximately 36,000 km from the earth's surface, it appears to stay in one spot in the sky, over the equator, 24 h a day. In reality, the satellite is moving around the earth but at the same angular speed that the earth is rotating, giving the illusion that it is hovering in the sky.

This is very useful. For instance, a satellite TV service can install home antennas that simply point to the spot in the sky where the satellite is located. Alternatively, a geostationary satellite can broadcast a signal to a large area (its "footprint") 24 h a day. Some geostationary satellite orbital locations are more economically preferable than others, depending on which regions of the earth are under the location.

A typical geostationary satellite will have several dozen transponders (relay amplifiers), each with a bandwidth of tens of MHz [149]. Such a satellite may weigh several thousand kilograms and consume several kilowatts using solar panels.

The number of microwave frequency bands used has increased over the years as the lower bands have become crowded and technology has improved. Frequency bands include L (1.5/1.6 GHz), S (1.9/2.2 GHz), C (4/6 GHz), Ku (11/14 GHz), and Ka (20/30 GHz) bands. Here the first number is the downlink band and the second number is the uplink band. The actual bandwidth of a signal may vary from about 15 MHz in the L band to several GHz in the Ka band [149].

It should be noted that extensive studies of satellite signal propagation under different weather and atmospheric conditions have been conducted. Excess power for overcoming rain attenuation is often budgeted above 11 GHz.

1.2.5.2 Low Earth Orbit Satellites

A more recent architecture is that of low earth orbit satellites. The most famous such system was Iridium from Motorola. It received its name because the original proposed 77 satellite network has the same number of satellites as the atomic number of the element Iridium. In fact, the actual system orbited had 66 satellites but the system name Iridium was kept.

The purpose of Iridium was to provide a global cell phone service. One would be able to use an Iridium phone anywhere in the world (even on the ocean or in the Arctic). Unfortunately, after spending five billion dollars to deploy the system, talking on Iridium cost a dollar or more a minute while local terrestrial cell phone service was under 25 cents a minute. While an effort was made to appeal to business travelers, the system was not profitable and was sold and is now operated by a private company.

Technologically though, the Iridium system is interesting. There are eleven satellites in each of six polar orbits (passing over the North Pole, south to the South Pole, and backup to the North Pole, see Fig. 1.1).

At any given time, several satellites are moving across the sky over any location on earth. Using several dozen spot beams, the system can support almost a quarter of a million conversations. Calls can be relayed from satellite to satellite.

It should be noted that when Iridium was hot, several competitors were proposed but not built. One used a “bent pipe” architecture where a call to a satellite would be beamed down from the same satellite to a ground station and then sent over the terrestrial phone network rather than being relayed from satellite to satellite. This was done in an effort to lower costs and simplify the design.

1.2.5.3 Other Orbits: MEO and Molniya

There are other possible orbits for communications satellites besides the geosynchronous and low earth orbits (Wikipedia).

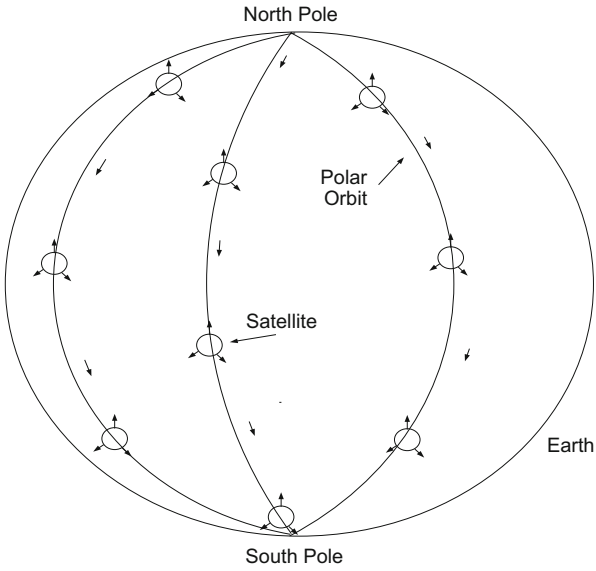


Fig. 1.1 Low earth orbit satellites (LEOS) in polar orbits

A satellite in *medium earth orbit (MEO)* is between 2000 and 35,786 km above the earth. Satellites in MEO operate in a fashion similar to LEO satellites but are above a spot on the earth's surface for longer periods than LEO satellites (normally from 2 to 8 h for MEOS). Because of the larger staying power and larger footprint (i.e., the area of earth coverage under a satellite) of MEO satellites, fewer are needed to form a global network. However, signal delays between the earth's surface and a MEO satellite is larger and signals are weaker (unless more power is used) than for a LEO satellite.

As one moves further north (or south), geostationary satellites which are over the equator appear lower in the sky (that is near the horizon). This creates problems of multipath interference as signals bounce off the ground so that multiple time shifted signals reach ground receivers. This creates a need for larger signal power. These signal power problems can be mitigated by launching satellites in *Molniya orbit* for countries such as Russia. Such orbits have the satellites spend a good portion of time over far northern latitudes with the satellite signal footprint moving minimally. A Molniya orbit satellite orbits the earth twice a day, being accessible over a spot in the far northern latitudes for 6–9 h every second orbit. Thus three Molniya orbit satellites can provide continuous connectivity for the intended region.

There are other issues with Molniya orbit satellites. Less energy is needed to place a satellite in a Molniya orbit than a geostationary orbit. Steerable antenna are needed to lock onto the satellite signal. Molniya satellites will also move through the Van Allen radiation belt four times a day.

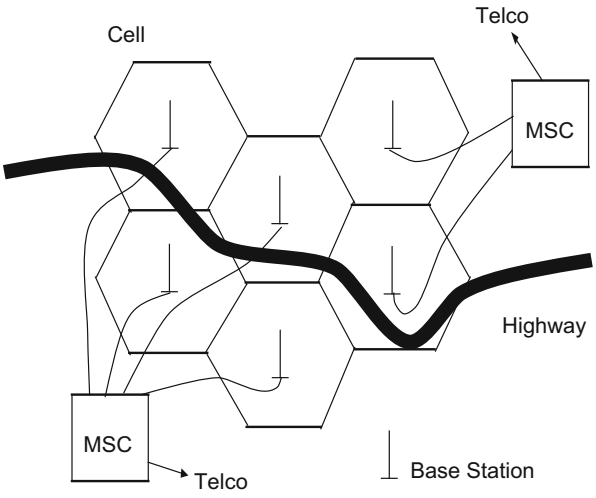


Fig. 1.2 Part of a cellular network

1.2.6 Cellular Systems

Starting around the early 1980s, cellular telephone systems which provide connectivity between mobile phones and the public switched telephone network were deployed. In such systems, signals go from/to a cell phone to/from a local "base station" antenna which is hard wired into the public switched telephone network. Figure 1.2 illustrates such a system. A geographic region such as a city or suburb is divided into geographic sub-regions called "cells."

In Fig. 1.2 base stations are shown at the center of cells. Nearby base stations are wired into a switching computer (the mobile switching center or MSC) that provides a path to the telephone network.

A cell phone making a call connects to the nearest base station (i.e., the base station with the strongest signal). Base stations and cell phones measure and communicate received power levels. If one is driving and one approaches a new base station, its signal will at some point become stronger than that of the original base station one is connected to and the system will then perform a "handoff." In a handoff, connectivity is changed from one base station to an adjacent one. Handoffs are transparent, the talking user is not aware when one occurs.

Calls to a cell phone involve a paging like mechanism that activates (rings) the called user's phone.

The first cellular system was deployed in 1979 in Japan by NTT. The first US cellular system was AMPS (Advanced Mobile Phone System) from AT&T. It was first deployed in 1983. These were first generation analog systems. Second generation systems (first deployed in 1991) were digital. The most popular is the European originated GSM (Global System for Mobile), which has been installed all

over the world. Third and fourth generation cellular systems provide increased data rates for such applications as Internet browsing, picture transmission, and streaming. Third generation cellular systems started with the WCDMA standard in Japan in 2001. Fourth generation cellular systems promised increases in data rate up to a factor of ten. The first fourth generation systems were WIMAX in North America and LTE, first offered in Scandinavia (Wikipedia).

1.2.7 Ad Hoc Networks

Ad hoc networks [99, 114] are radio networks where (often mobile) nodes can come together, transparently form a network without any user interaction and maintain the network as long as the nodes are in range of each other and energy supplies last [89, 118]. In an ad hoc network messages hop from node to node to reach an ultimate destination. For this reason ad hoc networks used to be called multi-hop radio networks. In fact, because of the non-linear dependence of energy on transmission distance, the use of several small hops uses much less energy than a single large hop, often by orders of magnitude.

Ad hoc network characteristics include multi-hop transmission, possibly mobility, and possibly limited energy to power the network nodes. Applications include mobile networks, emergency networks, wireless sensor networks, and ad hoc gatherings of people, as at a convention center.

Routing is an important issue for ad hoc networks. Two major categories of routing algorithms are topology based routing and position based routing. Topology based routing uses information on current links to perform the routing. Position based routing makes use of a knowledge of the geographic location of each node to route. The position information may be acquired from a service such as the Global Positioning System (GPS).

Topology based algorithms may be further divided into proactive and reactive algorithms. Proactive algorithms use information on current paths as inputs to classical routing algorithms. However, to keep this information current a large amount of control message traffic is needed, even if a path is unused. This overhead problem is exacerbated if there are many topology changes (say due to movement of the nodes).

On the other hand, reactive algorithms such as DSR, TORA, and AODV maintain routes only for paths currently in use to keep the amount of information and control overhead more manageable. Still, more control traffic is generated if there are many topology changes.

Position based routing does not require maintenance of routes, routing tables, or generation of large amounts of control traffic other than information regarding positions. “Geocasting” to a specific area can be simply implemented. A number of heuristics can be used in implementing position based routing.

1.2.8 Wireless Sensor Networks

The integration of wireless, computer, and sensor technology has the potential to make possible networks of miniature elements that can acquire sensor data and transmit the data to a human observer. Wireless sensor networks (WSN) are also known as wireless sensor and actuator networks (WSAN) when there is control of the sensors through bi-directional links (Wikipedia). Wireless sensor networks have received attention from researchers in universities, government, and industry because of their promise to become a revolutionary technology and the technical challenges that must be overcome to make this a reality. It is assumed that such wireless sensor networks will use ad hoc radio networks to forward data in a multi-hop mode of operation.

Typical parameters for a wireless sensor unit (including computation and networking circuitry) include a size from 1 mm to 1 cm, a weight less than 100 g, cost less than one dollar, and power consumption less than 100 μW [136]. By way of contrast, a wireless personal area network Bluetooth transceiver consumes more than a 1000 μW. A cubic millimeter wireless sensor can store, with battery technology, 1 Joule allowing a 10 μW energy consumption for 1 day [60]. Thus energy scavenging from light or vibration has been proposed. Note also that data rates are often relatively low for sensor data (100's bps to 100 Kbps).

Naturally, with these parameters, minimizing energy usage in wireless sensor networks becomes important. While in some applications wireless sensor networks may be needed for a day or less, there are many applications where a continuous source of power is necessary. Moreover, communication is much more energy expensive than computation. The energy cost of transmitting 1000 bits for 100 m equals the energy cost of processing three million instructions on a 100 million instructions per second/W processor (Wikipedia).

While military applications of wireless sensor networks are fairly obvious, there are many potential scientific and civilian applications of wireless sensor networks. Scientific applications include geophysical, environmental, and planetary exploration. One can imagine wireless sensor networks being used to locate forest fires, investigate volcanoes, measure weather, check water quality, monitor beach pollution, or record planetary surface conditions.

Biomedical applications include applications such as glucose level monitoring and retinal prosthesis [133]. Such applications are particularly demanding in terms of manufacturing sensors that can survive in and not affect the human body.

Sensors can be placed in machines (where vibration can sometimes supply energy) such as rotating machines, semiconductor processing chambers, robots, and engines. Wireless sensors in engines could be used for pollution control telemetry.

Finally, among many potential applications, wireless sensors could be placed in homes and buildings for climate control. Note that wiring a single sensor in a building can cost several hundred dollars. Ultimately, wireless sensors could be embedded in building materials.

1.3 Multiplexing

Multiplexing involves sending multiple signals over a single medium. Thomas Edison invented a four to one telegraph multiplexer that allowed four telegraph signals to be sent over one wire. The major forms of multiplexing for networking today are frequency division multiplexing (FDM), time division multiplexing (TDM), and spread spectrum. Each is now reviewed.

1.3.1 Frequency Division Multiplexing

Here a portion of spectrum (i.e., band of frequencies) is reserved for each channel (Fig. 1.3a). All channels are transmitted simultaneously but a tunable filter at the receiver only allows one channel at a time to be received. This is how AM, FM, and analog television signals are transmitted. Moreover, it is how distinct optical signals are transmitted over a single fiber using wavelength division multiplexing (WDM) technology.

1.3.2 Time Division Multiplexing

Time division multiplexing is a digital technology that, on a serial link, breaks time into equi-duration slots (Fig. 1.3b). A slot may hold a voice sample in a telephone

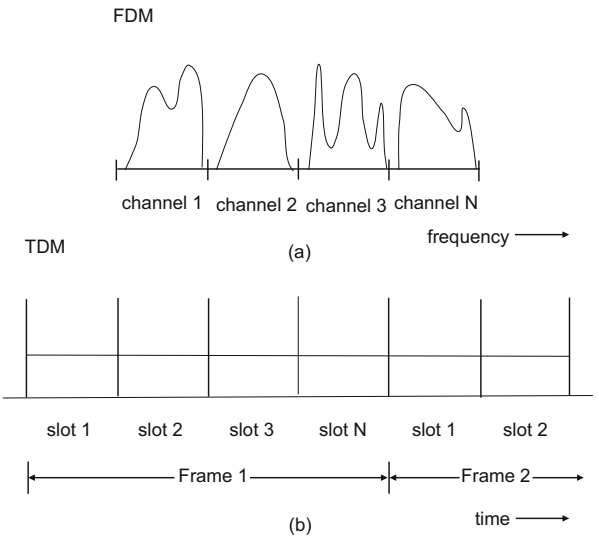


Fig. 1.3 (**a**) Frequency division multiplexing, (**b**) time division multiplexing

system or a packet in a packet switching system. A frame consists of N slots. Frames, and thus slots, repeat. A telephone channel might use slot 14 of 24 slots in a frame during the duration of a call, for instance.

Time division multiplexing is used in the second generation cellular system, GSM. It is also used in digital telephone switches. Such switches in fact use electronic devices called time slot interchangers that transfer voice samples from one slot to another to accomplish switching.

1.3.3 Frequency Hopping

Frequency hopping is one form of spread spectrum technology and is typically used on radio channels. The carrier (center) frequency of a transmission is pseudo-randomly hopped among a number of frequencies (Fig. 1.4a). The hopping is done in a deterministic, but random looking pattern that is known to both transmitter and receiver (i.e., "pseudo-random sequence"). If the hopping pattern is known only to the transmitter and receiver, one has good security. Frequency hopping also provides good interference rejection. Multiple transmissions can be multiplexed in the same local region if each uses a sufficiently different hopping pattern. Frequency hopping dates back to the era of World War II.

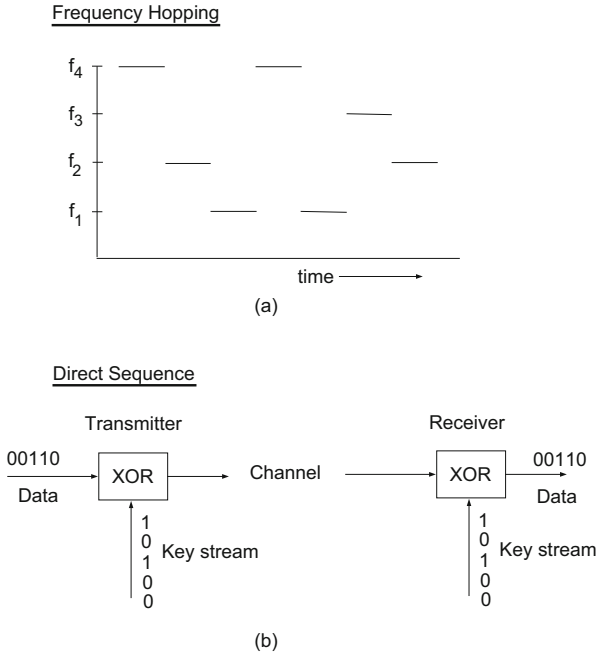


Fig. 1.4 (**a**) Frequency hopping spread spectrum, (**b**) direct sequence spread spectrum

Table 1.1 XOR truth table

Key	Data	Output
0	0	0
0	1	1
1	0	1
1	1	0

1.3.4 Direct Sequence Spread Spectrum

This alternative spread spectrum technology uses exclusive or (xor) gates as scramblers and de-scramblers (Fig. 1.4b). At the transmitter data is fed into one input of an xor gate and a pseudo-random key stream into the other input.

From the xor truth table (Table 1.1), one can see that if the key bit is a zero, the output bit equals the data bit. If the key bit is a one, the output bit is the complement of the data bit (0 becomes 1, 1 becomes 0). This scrambling action is quite strong under the proper conditions. Unscrambling can be performed by an xor gate at the receiver. The transmitter and receiver must use the same (synchronized) key stream for this to work. Again, multiple transmissions can be multiplexed in a local region if the key streams used for each transmission are sufficiently different.

1.4 Circuit Switching Versus Packet Switching

Two major architectures for networking and telecommunications are circuit switching and packet switching. Circuit switching is the older technology, going back to the years following the invention of the telephone in the late 1800s. As illustrated in Fig. 1.5a, for a telephone network, when a call has to be made from node A to node Z, a physical path with appropriate resources called a "circuit" is established. Resources include link bandwidth and switching resources. Establishing a circuit requires some set-up time before actual communication commences. Even if one momentarily stops talking, the circuit is still in operation. When the call is finished, link and switching resources are released for use by other calls. If insufficient resources are available to set up a call, the call is said to be blocked.

Packet switching was created during the 1960s. A packet is a group of bits consisting of header bits and payload bits. The header contains the source and destination address, priority levels, error check bits, and any other information that is needed. The payload is the actual information (data) to be transported. However, many packet switching systems have a maximum packet size. Thus, larger transmissions are split into many packets and the transmission is reconstituted at the receiver.

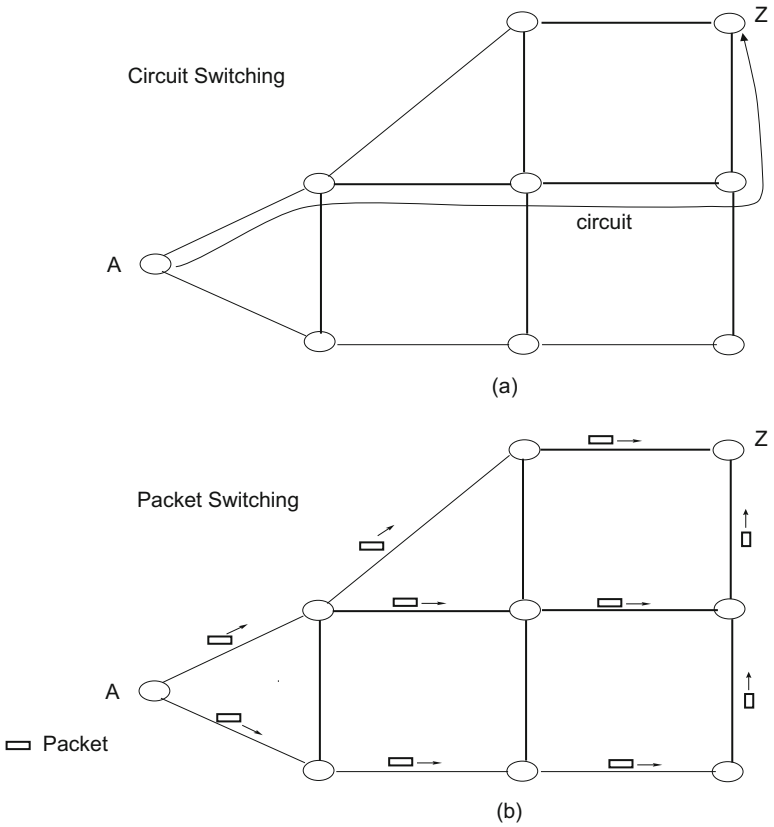


Fig. 1.5 (**a**) Circuit switching, (**b**) packet switching

The diagram of Fig. 1.5b shows packets, possibly from the same transmission, taking multiple routes from node A to node Z. This is called datagram or connectionless oriented service. Packets may indeed take different routes in this type of service as nodal routing tables are updated periodically in the middle of a transmission.

A hybrid type of service is the use of "virtual circuits" or connection oriented service. Here packets belonging to the same transmission are forced to take the same serial path through the network. A virtual circuit has an identification number which is placed in packet headers to be used by nodes to continue forwarding the packet along its preset (circuit) path. As in circuit switching, a virtual circuit needs to be set up prior to its use for communication. That is, entries need to be made in routing tables implementing the virtual circuit.

An advantage of virtual circuit usage is that packets arrive at the destination in the same order that they were sent. This avoids the need for buffers for reassembling transmissions (reassembly buffers) that are needed when packets arriving at the destination are not in order, as in datagram service.

Packet switching is advantageous when traffic is bursty (occurs at irregular intervals) and individual transmissions are short. It is a very efficient way of sharing network resources when there are many such transmissions. Circuit switching is not well suited for bursty and short transmissions. It is more efficacious when transmissions are relatively long (to minimize set-up time overhead) and provide a constant traffic rate (to well utilize the dedicated circuit resource).

1.5 Layered Protocols

Protocols are the rules of operation of a network. A common way to engineer a complex system is to break it into more manageable and coherent components. Network protocols are often divided into layers in the layered protocol approach. Figure 1.6 illustrates the generic OSI (open systems interconnection) protocol stack. Proprietary protocols may have different names for the layers and/or a different layer organization but pretty much all networking protocols have the same functionality.

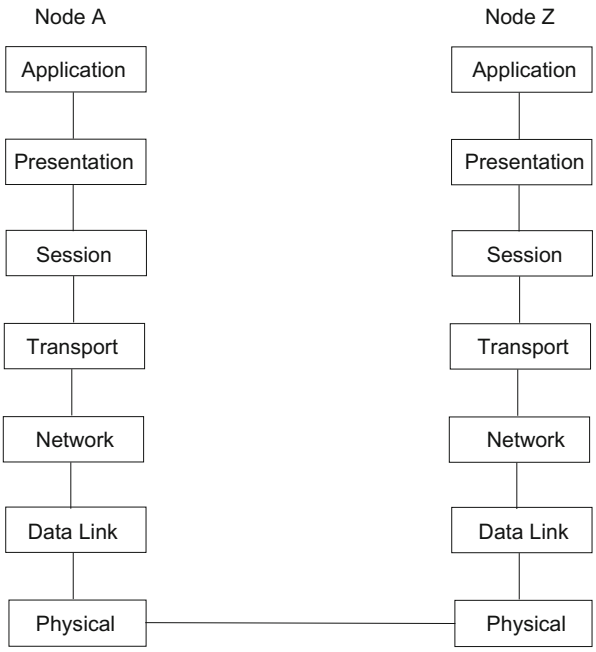


Fig. 1.6 OSI protocol stack for a communicating source and destination

Transmissions in a layered architecture (see Fig. 1.6) move from the source's top layer (application), down the stack to the physical layer, through a physical channel in a network, to the destination's physical layer, up the destination stack to the destination application layer. Note that any communication between peer layers must move down one stack, across and up the receiver's stack. It should also be noted that if a transmission passes through an intermediate node, only some lower layers (e.g., network, data link, and physical) may be used at the intermediate nodes.

It is interesting that a packet moving down the source's stack may have its header grow as each layer may append information to the header. At the destination, each layer may remove information from the packet header, causing it to decrease in size as it moves up the stack.

In a particular implementation, some layers may be larger and more complex while others are relatively simple.

In the following, we briefly discuss each layer.

1.5.1 Application Layer

Applications for networking include email, remote login, file transfer, and the world wide web. But an application may also be more specialized, such as distributed software to run a network of catalog company order depots.

1.5.2 Presentation Layer

This layer controls how information is formatted, such as on a screen (number of lines, number of characters across).

1.5.3 Session Layer

This layer is important for managing a session, as in remote logins. In other cases, this is not a concern.

1.5.4 Transport Layer

This layer can be thought of as an interface between the upper and lower layers. More importantly, it is designed to give the impression to the layers above that they are dealing with a reliable network, even though the layers below the transport layer may not be perfectly reliable. For this reason, some think of the transport layer as the most important layer.

1.5.5 Network Layer

The network layer manages multiple links. Its most important function is to do routing. Routing involves selecting the best path for a circuit or packet stream.

1.5.6 Data Link Layer

Whereas, the network layer manages multiple link functions, a data link protocol manages a single link. One of its potential functions is encryption, which can either be done on a link by link basis (i.e., at the data link layer) or on an end-to-end basis (i.e., at the transport layer) or both. End-to-end encryption is a more conservative choice as one is never sure what type of sub-network a transmission may pass through and what its level of encryption, if any, is.

1.5.7 Physical Layer

The physical layer is concerned with the raw transmission of bits. Thus, it includes engineering physical transmission media, modulation and demodulation, and radio technology. Many communication engineers work on physical layer aspects of networks. Again, the physical layer of a protocol stack is the only layer that provides actual direct connectivity to peer layers.

Introductory texts on networking usually discuss layered protocols in detail.

Chapter 2
Ethernet

2.1 Introduction

Local area networks (LANs) are networks that cover a small area as in a department, in a company, or university. In the early 1980s, the three major local area networks were Ethernet (IEEE standard 802.3), Token Ring (802.5 and used extensively by IBM), and Token Bus (802.4, intended for manufacturing plants). However, over the years, Ethernet [149] has become the most popular wired local area network standard. While maintaining a low cost, it has gone through six versions, most ten times faster than the previous version (10 Mbps, 100 Mbps, 1 Gbps, 10 Gbps, 40 Gbps, 100 Gbps, and in the works 400 Gbps).

Ethernet was invented at the Xerox Palo Alto Research Center (PARC) by Metcalfe and Boggs [92]. It is similar in spirit to the earlier Aloha radio protocol, though the scale is smaller. IEEE's 802.3 committee produced the first Ethernet standard. Xerox never produced Ethernet commercially but other companies did.

In going from one Ethernet version to the next, the IEEE 802.3 committee sought to make each version similar to the previous ones and to use existing technology. In the following, we now discuss the various versions of Ethernet.

2.2 10 Mbps Ethernet

Back in the 1980s, Ethernet was originally wired using coaxial cable. As in Fig. 2.1a, a coaxial cable was snaked through the floor or ceiling and computers attached to it along its length. The coaxial cable acted as a private radio channel that each computer would monitor. If a station had a packet to send, it would send it

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_2

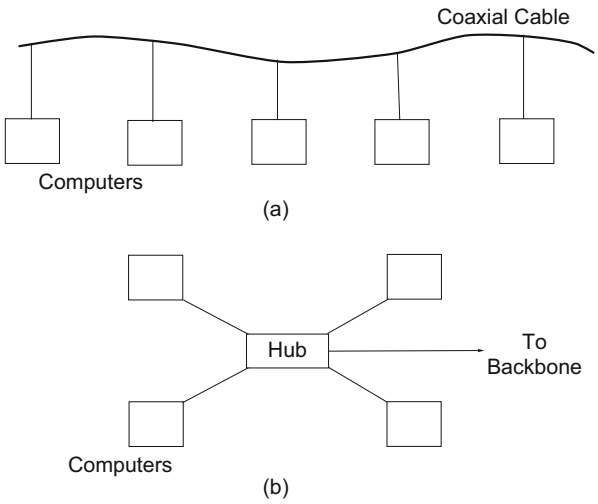


Fig. 2.1 Ethernet wiring using (**a**) coaxial cable and (**b**) hub topology

immediately if the channel was idle. If the station sensed the channel to be busy, it would wait until the channel was free. In all of this, only one transmission can be on the channel at one time.

A problem occurs if two or more stations sense the channel to be idle at about the same time and attempt to transmit simultaneously. The packets overlap in the cable and are garbled. This is a collision. The stations involved, using analog electronics, can detect the collision, stop transmitting, and reschedule their transmissions.

Thus, the price one pays for this completely decentralized access protocol is the presence of utilization lowering collisions. The protocol used goes by the name 1-persistent CSMA/CD (Carrier Sense Multiple Access with Collision Detection). The name is pretty much self-explanatory except that 1-persistent refers to the fact that a station with a packet to send attempts this on an idle channel with a probability of 1.0. In a CSMA/CD protocol, if the bit rate is 10 Mbps, the actual useful information transport can be significantly less because of collisions (or occasional idleness).

In the case of a collision, the rescheduling algorithm used is called Binary Exponential Backoff. Under this protocol, two or more stations experiencing a collision randomly reschedule over a time window with a default of 51 μs for a 500 m network. If a station becomes involved in a second collision, it doubles its time window size and attempts again to randomly reschedule its transmission. Windows may be doubled in size up to ten times. Once a packet is successfully transmitted, the time window size drops back to the default (smallest) value for that packet's station. Thus, this protocol at a station has no long term memory regarding past transmissions.

Table 2.1 Ethernet frame format

Field	Length
Preamble	7 bytes
Frame delimiter	1 byte
Destination address	2 or 6 bytes
Source address	2 or 6 bytes
Data length	2 bytes
Data	up to 1500 bytes
Pad	variable
CRC Checksum	4 bytes

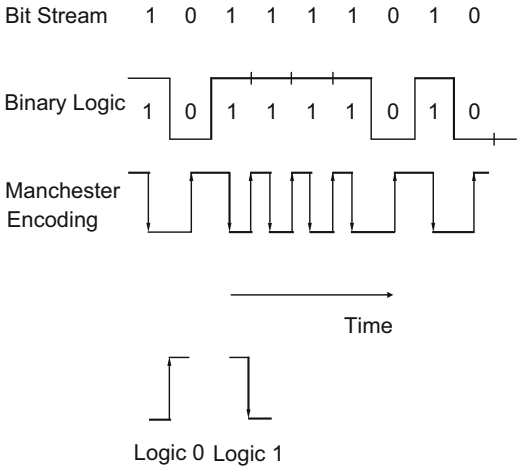


Fig. 2.2 Manchester encoding

Table 2.1 above shows the fields in the 10 Mbp Ethernet frame. A frame is the name for a packet at the data link layer. The preamble is for communication receiver synchronization purposes. Addresses are either local (2 bytes) or global (6 bytes). Note that Ethernet addresses are different from IP addresses. Different amounts of data can be accommodated up to 1500 bytes. Transmissions longer than 1500 bytes of data must be segmented into multiple packets. The pad field is used to guarantee that the frame is at least 64 bytes in length (minimum frame size) if the frame would be less than 64 bytes in length. Finally the checksum is based on CRC error detecting coding.

A problem with digital receivers is that they require many 0 to 1 and 1 to 0 transitions to properly lock onto a signal. But long runs of 1's or 0's are not uncommon in data. To provide many transitions between logic levels, even if the data has a long run of one logic level, 10 Mbps Ethernet uses Manchester encoding.

Referring to Fig. 2.2, under Manchester encoding, if a logic 0 needs to be sent, a transition is made for 0 to 1 (low to high voltage) and if a logic 1 needs to be sent, the opposite transition is made for 1 to 0 (high to low voltage). The voltage level makes a return to its original level at the end of a bit as necessary. Note that the "signaling rate" is variable. That is, the number of transitions per second is twice

Table 2.2 Original ethernet wiring

Cable	Type	Maximum size
10Base5	Thick coax	500 m
10Base2	Thin coax	200 m
10Base-T	Twisted pair	100 m
10Base-F	Fiber optics	2 km

the data rate for long runs of a logic level and is equal to the data rate if the logic level alternates. For this reason, Manchester encoding is said to have an efficiency of 50%. More modern signaling codes, such as 4B5B, achieve a higher efficiency (see Fast Ethernet below).

During the 1980s, Ethernets were wired with linear coaxial cables. Today hubs are commonly used (Fig. 2.2b). These are boxes (some smaller than a cigar box) that computers tie into, in a star type wiring pattern, with the hub at the center of the star.

A hub may internally have multiple cards, each of which has multiple external Ethernet connections. A high speed (in the gigabits) proprietary bus interconnects the cards. Cards may mimic a CSMA/CD Ethernet with collisions (shared hub) or use buffers at each input (switched hub). In a switched hub, multiple packets may be received simultaneously without collisions, raising throughput at the expense of some delay.

The next table (Table 2.2) illustrates Ethernet wiring. In "10 Base5", the 10 stands for 10 Mbps and the 5 for the 500 m maximum size. Used in the early 1980s, 10 Base5 used vampire taps which would puncture the cable. Also at the time, 10 Base2 used T junctions and BNC connectors as wiring hardware. Today, 10 Base-T is the most common wiring solution for 10 Mbps Ethernet. Fiber optics, 10 Base-F, was only intended for runs between buildings, but a higher data rate protocol would probably be used today for this purpose.

2.3 Fast Ethernet

As the original 10 Mbps Ethernet became popular and the years passed, traffic on Ethernet networks continued to grow. To maintain performance, network administrators were forced to segment Ethernet networks into smaller networks (each handling a smaller number of computers) connected by a spaghetti-like arrangement of inter-connecting repeaters, bridges, and routers. In 1992, IEEE assigned the 802.3 committee the task of developing a faster local area network protocol.

The committee agreed on a 100 Mbps protocol that would incorporate as much of the existing Ethernet protocol/technology as possible to gain acceptance and so that they could move quickly. The resulting protocol, IEEE 802.3u, was called Fast Ethernet.

Fast Ethernet is only implemented with hubs, in a star topology (Fig. 2.1b). There are three major wiring options (Table 2.3).

Table 2.3 Fast ethernet wiring

Cable	Type	Maximum Size
100Base-T4	Twisted pair	100 m
100Base-TX	Twisted pair	100 m
100Base-FX	Fiber optics	2 km

The original Ethernet has a data rate of 10 Mbps and a maximum signaling rate of 20 MHz (recall that the Manchester encoding used was 50% efficient). Fast Ethernet 100 Base-T4 with its data rate of 100 Mbps has a signaling speed of 25 MHz, not 200 MHz. How is this accomplished?

Fast Ethernet 100 Base-T4 actually uses four twisted pairs per cable. Three twisted pairs carry signals from its hub to a PC. Each of the three twisted pairs uses ternary (not binary) signaling using 3 logic levels. Thus, one of $3 \times 3 \times 3 = 27$ symbols can be sent at once. Only 16 symbols are used though, which is equivalent to sending 4 bits at once. With 25 MHz clocking 25 MHz $\times 4$ bits yields a data rate of 100 Mbps. The channel from the PC to hub operates at 33 MHz. For most PC applications, an asymmetrical connection with more capacity from hub to PC for downloads is acceptable. Category 3 or 5 unshielded twisted pair wiring is used for 100 Base-T4.

An alternative to 100 Base-T4 is 100 Base-TX. This uses two twisted pairs, with 100 Mbps in each direction. However, 100 Base-T4 has a signaling rate of only 125 MHz. It accomplishes this using 4B5B (Four Bit Five Bit) encoding rather than Manchester encoding. Under 4B5B, every four bits is mapped into five bits in such a way that there are many transitions for digital receivers to lock onto, irrespective of the actual data stream. Since four bits are mapped into five bits, 4B5B is 80% efficient. Thus, 125 MHz times 0.8 yields 100 Mbps.

Finally, 100 Base-FX uses two strands of the lower performing multi-mode fiber. It has 100 Mbps in both directions and is for runs (say between buildings) of up to 2 km.

It should be noted that Fast Ethernet uses the signaling method for twisted pair (for 100 Base-TX) and fiber (100 Base-FX) borrowed from FDDI. The FDDI protocol was a 100 Mbps token ring protocol used as a backbone in the 1980s.

To maintain channel efficiency (utilization) at 100 Mbps for CSMA/CD implementations, versus the original 10 Mbps, the maximum network size of Fast Ethernet is about ten times smaller than that of the original Ethernet. The trade-off can be seen in the Ethernet design equation [122].

2.4 Gigabit Ethernet

The ever growing amount of network traffic brought on by the growth of applications and more powerful computers motivated a revised, faster version of Ethernet. Approved in 1998, the next version of Ethernet operates at 1000 Mbps or 1 Gbps

and is known as Gigabit Ethernet, or 802.3z. As much as possible, the Ethernet committee sought to utilize existing Ethernet features.

Gigabit Ethernet wiring is either between two computers directly or, as is more common, in a star topology with a hub or switch in the center of the star. In this connection, it is appropriate to say something about the distinction between a hub and switch. A shared medium hub uses the established CSMA/CD protocol so collisions can occur. At most, one attached station can successfully transmit through the hub at a time, as one would expect with CSMA/CD. The half duplex Gigabit Ethernet mode uses shared medium hubs.

A "switch," on the other hand, does not use CSMA/CD. Rather, the use of buffers means multiple attached stations may send and receive distinct communications to/from the switch at the same time. The use of multiple simultaneous transmissions means that switch throughput is substantially greater than that of a single input line. Level 2 switches are usually implemented in software, level 3 switches implement routing functions in hardware [144]. Full duplex Gigabit Ethernet most often uses switches.

In terms of wiring, Gigabit Ethernet has two fiber optic options (1000 Base-SX and 1000 Base-LX), a copper option (1000 Base-CX) and a twisted pair option (1000-Base T).

The Gigabit Ethernet fiber option deserves some comment. It makes use of 8B10B encoding, which is similar in its operation to Fast Ethernet's 4B5B. Under 8B10B, eight bits (1 byte) are mapped into 10 bits. The extra redundancy this involves allows each 10 bits not to have an excessive number of bits of the same type in a row or too many bits of one type in each of 10 bits. Thus, there are sufficient transitions from 1 to 0 and 0 to 1 or the data stream even if the data has a long run of 1's and 0's.

Gigabit Ethernet using twisted pair uses five logic levels on each wire. Four of the logic levels convey data and the fifth is for control signaling. With four data logic levels, two bits are communicated at once or eight bits over all four wires at a time. Thus the signaling rate is 1 Gbps/8 or 125 MHz.

In terms of utilization under CSMA/CD operation, if the maximum segment size had been reduced by a factor of 10 as was done in going from the original Ethernet to Fast Ethernet, only very small gigabit networks could have been supported. To compensate for the ten times increase in data rate relative to Fast Ethernet, the minimum frame size for Gigabit Ethernet was increased (by a factor of eight) to 512 bytes from Fast Ethernet's 512 bits (see [122] for a discussion of the Ethernet equation that governs this).

Another technique that helps Gigabit Ethernet's efficiency is frame bursting. Under frame bursting, a series of frames are sent in a single burst.

Gigabit Ethernet's range is at least 500 m for most of the fiber options and about 200 m for twisted pair [144, 149].

2.5 10 Gigabit Ethernet

Considering the improvement in Ethernet data rate over the years, it is not too surprising that a 10 Gbps Ethernet was developed [143, 154]. Continuing the increases in data rate by a factor of ten that have characterized the Ethernet standards, 10 Gbps (or 10,000 Mbps) Ethernet is ten times faster than Gigabit Ethernet. Applications include backbones, campus size networks, and metropolitan and wide area networks. This latter application is aided by the fact that the 10 Gbps data rate is comparable with a basic SONET fiber optic transmission standard rate. See a later chapter for more information on SONET.

There are eight implementations of 10 Gbps Ethernet. It can use four transceiver types (one four wavelength parallel system and three serial systems with a number of multi-mode and single mode fiber options). Like earlier versions of Ethernet, it uses CRC error coding. It operates in full duplex non-CSMA/CD mode. It can go more than 40 km via single mode fiber.

To lower the speed at which the MAC (Media Access Control) layer processes the data stream, the MAC operates in parallel on four 2.5 Gbps streams (lanes). As illustrated in Fig. 2.3, bytes in an arriving 10 Gbps serial transmission are placed in parallel in the four lanes.

There is a 12 byte inter-packet gap (IPG) which is the minimum gap between packets. Normally, it would not be easy to predict the ending byte lane of the previous packet, so it would be difficult to determine the starting lane of the next transmission. The solution is to have a starting byte in a packet always occupy lane 0. The IPG is found using a pad (add in extra 1 to 3 bytes), a shrink (subtract 1 to 3 bytes), or through combination averaging (average of 12 bytes achieved through a combination of pads and shrinks). Note that padding introduces extra overhead in some implementations.

In terms of the protocol stack, this can be visualized as in Fig. 2.4.

The PCS, PMA, and PMD sublayers use parallel lanes for processing. In terms of the sublayers, they are:

Reconciliation: Command translator that maps terminology and commands in MAC into electrical format appropriate for physical layer.

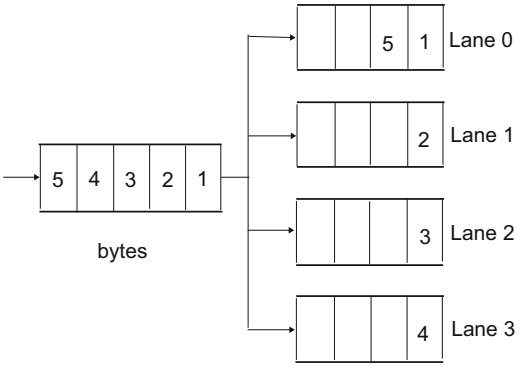


Fig. 2.3 Four parallel lanes for 10 Gigabit Ethernet

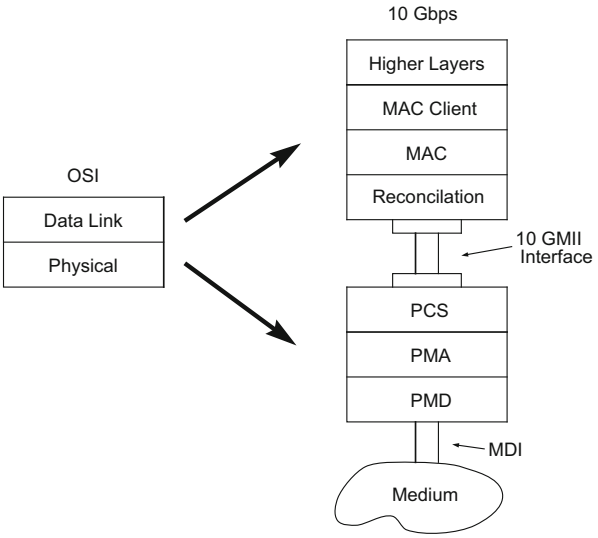


Fig. 2.4 Protocol stack for 10 Gbps Ethernet

PCS: Physical Coding Sublayer.
PMA: Physical Medium Attachment (at transmitter serialize code groups into bit stream, at receiver synchronization for data decoding).
PMD: Physical Medium Dependent (includes amplification, modulation, wave shaping).
MDI: Medium Dependent Interface (i.e., connector).

2.6 40/100 Gigabit Ethernet

Over the years Ethernet has been attractive to users because of its relatively low cost, robustness, and its ability to provide an interoperable network service. Users have also liked the wide vendor availability of Ethernet related products. However, even with the release of gigabit and 10 Gbps Ethernet demand for bandwidth continued to grow. Network equipment shipments can grow at a 17% a year rate. Internet traffic grows at 75–125% a year. Computer performance doubles every 24 months. A 2008 projection was that within 4 years 40 Gbps would be needed.

One of the applications driving this growth is the increasing use of data centers (see the latter chapter on this topic). These facilities house server farms for hosting web services and cloud computing services. Projections indicate a need for 100 Gbps of data transfer capacity from switch to switch. Also 100 Gbps will have applications between buildings, within campuses, and for metropolitan area networks (MAN) and wide area networks (WAN).

In July 2006 a committee was convened to explore increasing the data rate of Ethernet beyond 10 Gbps. In 2010 standards for 40 Gbps and 100 Gbps Ethernet were approved. This discussion is based on [35, 101].

2.6.1 40/100 Gigabit Technology

In implementing 40 and 100 gigabit Ethernet some of the objectives are:

- MAC (medium access control) data rates of 40 and 100 gigabit per second.
- Full duplex is only supported (i.e., two way communication).
- Maintain the existing minimum and maximum frame length.
- Use the current frame format and MAC layer.
- Optical transport network (OTN) support.

A variety of transmission media can carry 40 and 100 gigabit Ethernet as Table 2.4 illustrates.

Figure 2.5 illustrates the protocol stack for 40 and 100 gigabit Ethernet. In the figure one has the physical coding sublayer (PCS), the forward error correction

Table 2.4 40/100 Gbps Ethernet

40 Gbps	100 Gbps
≥10 km single mode fiber	≥ 40 km single mode fiber
≥100 m multi-mode fiber	≥ 10 km single mode fiber
≥10 m copper cable	≥ 100 m multi-mode fiber
≥1 m backplane	≥ 10 m copper cable

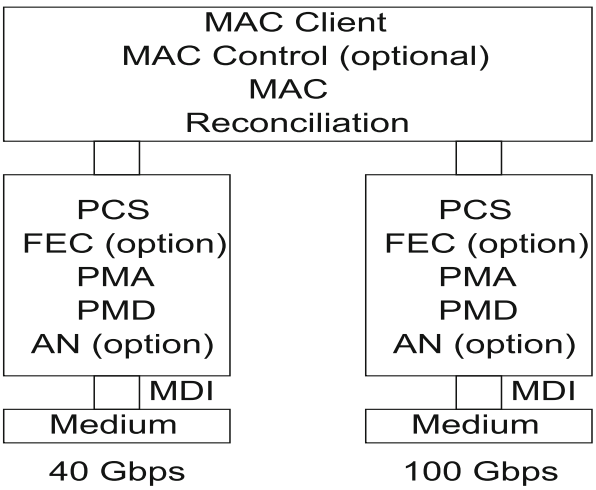


Fig. 2.5 40 and 100 Gbps Ethernet protocol functions

sublayer (FEC), physical medium attachment sublayer (PMA), physical medium dependent sublayer (PMD), and the auto-negotiation sublayer (AN). Here also MDI is the medium dependent interface or the connector.

In the physical coding sublayer 64B/66B coding is used, mapping 64 bits into 66 bits to provide enough transitions between 0 and 1 for digital receivers. As in 10 gigabit Ethernet, the concept of parallel lanes is used in 40 and 100 gigabit Ethernet. A 66 bit block is distributed in round robin fashion on the PCS lanes. Specifically for 40 gigabit Ethernet there are 4 PCS lanes that support 1, 2, or 4 channels or wavelengths. For 100 gigabit Ethernet there are 20 PCS lanes that support 1, 2, 4, 5, 10, or 20 channels or wavelengths. As an example, a 100 gigabit Ethernet may use 5 parallel wavelengths over a fiber, each carrying 20 Gbps.

Created PCS lanes can be multiplexed into any interface width that is supported. There is a unique lane marker for each PCS lane which is inserted periodically. Bandwidth for the lane marker comes from periodically deleting the inter-packet gap (IPG) in a lane. All bits in the same lane follow the same physical path no matter how multiplexing is done.

The receiver reassembles the PCS lanes by demultiplexing bits and also realigns the PCS lanes taking into account the skewness of the lanes. Advantages of this include the fact that all encoding, deskew, and scrambling functions are implemented on a CMOS device located on the host and there is minimal bit processing except for using an optical module for multiplexing.

Finally, clocking takes place at 1/64th of the data rate (625 MHz for 40 gigabits and 1.5625 GHz for 100 gigabits). More information on 40 and 100 gigabit Ethernet can be found on www.ethernetalliance.org.

2.7 Higher Ethernet Speeds

2.7.1 Introduction

In 2012 the IEEE 802.3 Ethernet committee did a data rate growth assessment and came to the following conclusions [36]:

- The need for increased data rates was increasing more rapidly for "network aggregation nodes," than for end-station applications. Network aggregation nodes combine traffic from very many end users. Perhaps the most important example of this is data centers.
- The compound annual growth rate (CAGR) that should be supported is 58%. The biggest growth rates were in data intensive science and in the financial industry.
- In 2012 the 802.3 committee projected a need in data rate capacities of 1 Tbps (i.e., 10^{12} bps) by 2015 and 10 Tbps by 2020.

As John D'Ambrosia and Paul Mooney put it in a 2013 white paper [36]:

"Bandwidth growth is unrelenting everywhere across Ethernet networking. Every day, *more* users are *more* quickly accessing the Internet in *more* ways, to utilize *more* applications and consume *more* content that demands *more* bandwidth every day. More, more, more…"

In fact there is a cyclic action where enabling higher data rates enables applications which creates a need for increased data rates leading to enabling higher data rates and on and on… [36].

2.7.2 The Road to Higher Speeds

The initial thought of the Ethernet community was that it was time to create a 400 Gbps version of Ethernet. The IEEE P802.3bs 400 Gigabit Ethernet (400 GbE) committee (known as the 400 Gbps Ethernet study group) first met in May 2014. One might ask why not go for a ten fold increase of speed from 100 Gbps Ethernet to 1 Tbps Ethernet? The feeling was that a 400 Gbps implementation was doable with existing technology whereas a 1 Tbps implementation would involve new technology and a need for more research and development.

As of 2016 there were a number of parallel efforts going on under the aegis of IEEE standards bodies. These include a 200 Gbps version of Ethernet as well as a 400 Gbps version. Moreover higher data rates are built out of parallel lanes or channels. In this light, to be started is work on a 50 Gbps version of Ethernet which can be used as a building block for higher rates.

Both multi-mode and single mode fiber versions of higher data rate Ethernets are being considered.

In terms of goals for higher rates, one has [37] (http://www.wikipedia.org):

- Full duplex operation only (as has been true since 10 Gbps Ethernet).
- MAC data rates of 200 and 400 Gbps.
- Use existing Ethernet frame format using the Ethernet MAC.
- Keep the minimum and maximum frame size of the existing standard.
- The bit error rate should be 10^{-13} or the frame loss equivalent. Note that for 10, 40 and 100 Gbps Ethernet versions the bit error rate was 10^{-12}.
- OTN (optical transport network for Ethernet) should be supported. Energy-Efficient Ethernet (EEE) is optionally supported.
- Optional 400 GBps attachment unit interfaces for chip-to-chip and chip-to-module applications.
- The physical layer supports link distances in Table 2.5.

While the implementation details need to be filled in, the trend for Ethernet is a march to higher and higher data rates.

Table 2.5 200/400 Gbps Ethernet

200 Gbps	400 Gbps
≥500 m single mode fiber	≥ 100 m multi-mode fiber
≥2 km single mode fiber	≥ 500 m single mode fiber
≥10 km single mode fiber	≥ 2 km single mode fiber
	≥10 km single mode fiber

2.8 Conclusion

For 35 years Ethernet has continually transformed itself by way of higher data rates to meet increasing demand for networking services. It will be interesting to see what the future holds.

Chapter 3
InfiniBand

3.1 Introduction

InfiniBand is a proprietary high speed, low latency networking technology that has been most successful as an interconnection technology in High Performance Computing (HPC) centers. High performance computing involves powerful (often clustered) computers usually used for big scientific and technical computing problems. As of 2015, more than 50 percent of the world's such supercomputers on the Top 500 list made use of InfiniBand [164]. InfiniBand is also a contender as an interconnect for other large data processing centers but has been slower to gain acceptance there partly because of perceptions of higher cost and the ubiquity and familiarity of Ethernet interconnects.

InfiniBand "lanes" can be put in parallel to boost throughput. For instance, circa 2012 [100], one could have a 1× 14 Gbps lane, or 56 Gbps across four lanes or 168 Gbps over 12 lanes. Circa 2015 one can have (EDR: Enhanced Data Rate) 25 Gbps for 1 lane, 100 Gbps for 4 lanes, or 300 Gbps for 12 lanes. InfiniBand data rates usually come to market at least a year before similar Ethernet data rates [94].

In 2015 the InfiniBand Trade Association released Release 1.3 of Volume 1 of the InfiniBand Architecture Specification. Volume 1 presents the "core" of the InfiniBand Architecture. It describes requirements for InfiniBand switches, routers, host channel adapters, and target channel adapters [100]. Volume 1 describes the architecture for managing an InfiniBand fabric. Volume 2 describes the architecture's physical aspects.

Release 1.3 of Volume 1 enables network administrators to "more easily install, maintain and optimize very large InfiniBand clusters" [167]. This includes better visibility into the switch hierarchy, better and faster diagnostics for connectivity problems, and better network statistics.

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_3

Some of the following discussion is based on the more extensive treatment in [50]. InfiniBand started to be offered in 1999. A key feature is that often InfiniBand transfers data directly from the memory of one computer to the memory of another without going through the computer's operating systems. This is referred to as "RDMA" or Remote Direct Memory Access. This is an extension of Direct Memory Access (DMA) used in PCs. In DMA a DMA engine (controller) allows memory access without involving the CPU processor. Off-loading this function to the DMA engine makes for a more efficient use of the CPU. The difference between RDMA and DMA is that RDMA is done between remote (separate or distant) machines whereas DMA is done on a single machine.

One of the simplifying features of InfiniBand is that it provides a "messaging service" that applications can directly access. This compares to TCP/IP over Ethernet, which is byte oriented rather than message oriented.

3.2 A First Look

The message service of InfiniBand is easy to use. It can allow communication from an application to other applications or processes or to gain access to storage. In using the messaging service the operating system is not needed. Instead an application directly accesses the messaging service rather than one of the server's communication resources.

InfiniBand "creates" a channel application. Applications making use of the service can either be kernel[1] applications (such as file systems) or in user space. All of InfiniBand is geared towards supporting this top-down messaging service. Channels serve as pipes (i.e., connections) between disjoint virtual address spaces. These could also be disjoint physical address spaces (that is distinct servers separated by distance).

3.2.1 Queue Pairs

The endpoints of a channel are the send and receive queues. This is also known as a queue pair. When an application requires one or more connections more Queue Pairs (QP's) are generated. The Queue Pair maps directly into the virtual address spaces of each application. This idea is called "channel I/O."

[1] A kernel is the most basic part of most operating systems. It connects applications to data processing in the computer's hardware. Kernel address space is used exclusively by the kernel, kernel extensions, and most device drivers while user space is where the user applications reside and work.

3.2.2 Transfer Semantics

There are two ways in which data can be transferred in InfiniBand:

- SEND/RECEIVE: The application on the receiver side provides a data structure for received messages. The data structure is pre-posted on the receiver queue. Actually the sending side doesn't "see" the buffers or the data structure on the receiver side. The sending side just SENDS one or more messages and the receiver side RECEIVES them.
- RDMA READ/RDMA WRITE: The steps are as follows:

  (a) A buffer is registered in the receiver side application's virtual address space by the receiver side application.
  (b) Control of the buffer is passed to the sending side by the receiver.
  (c) The sending side uses the RDMA READ or RDMA WRITE operations to either read or write data in that buffer.

3.2.3 InfiniBand Verbs

InfiniBand "verbs", is an abstraction of the functionality needed for RDMA applications. One can think of verbs as a low level abstraction for RDMA programming. The use of verbs yields the best performance in terms of latency, throughput, and message rate. Verbs can be thought of as building blocks for numerous applications such as sockets, storage, and parallel programming. Using levels of abstractions higher than verbs may result in reduced performance.

Verbs come in two major groups [17]:

Control Path: These manage resources and normally need a context switch. Examples include Create, Destroy, Modify, Query, and Work with Events.
Data Path: These utilize resources to send and receive data and normally doesn't need a context switch. Examples include Post Send, Post Receive, Poll CQ, and Request for Completion Event.

3.3 The InfiniBand Protocol

InfiniBand messages may be up to 2^{31} bytes. Messages are partitioned (segmented) into packets by the InfiniBand hardware. The packet size used is selected to make the best use of network bandwidth. InfiniBand switches and routers are used for transmitting packets through InfiniBand.

The generic OSI protocol stack layers that correspond to the InfiniBand messaging service are illustrated in Fig. 3.1. Here "SW" stands for software.

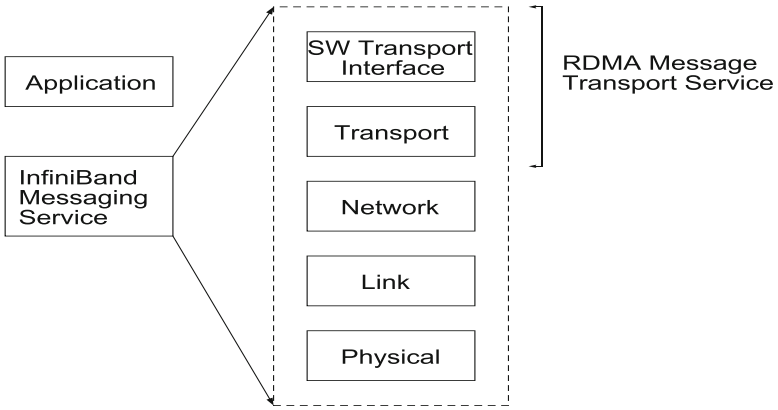


Fig. 3.1 InfiniBand equivalency to OSI protocol stack (after Grun [50])

An InfiniBand switch is similar in theory to other types of common switches but is adapted to InfiniBand performance and cost goals. InfiniBand switches use cut through switching for better performance. Under cut through switching a node can start forwarding a packet before it is completely received by the node. InfiniBand link layer flow control is employed so during standard operation packets are not dropped. Ethernet has more loss than InfiniBand.

Software for InfiniBand is made up of upper layer protocols (ULPs) and libraries. Mid-layer functions support the ULPs. There are hardware specific data drivers.

3.4 InfiniBand for HPC

The Message Passing Interface (MPI) is the leading standard and model for parallel system communication. Why use MPI? It offers a communication service to the distributed processes making up an HPC (High Performance Computing) application. In fact MPI middleware[2] is used by InfiniBand. The MPI middleware in InfiniBand is allowed to communicate between machines in a cluster without the involvement of the CPUs in the cluster. The copy avoidance[3] architecture and stack bypass feature of InfiniBand provide extremely low application to application delay (latency), high bandwidth, and low CPU loading.

[2]Middleware is software which connects application software to the operating system. An example of middleware is a web browser. Some types of middleware are eventually incorporated into newer versions of operating systems. An example of this migration is TCP/IP.

[3]Copy avoidance methods use less copying of data to memory by the operating system leading to higher data transfer rates. Zero copy methods make no copies.

Some other InfiniBand options, particularly for use with storage, include:

SDP (Socket Direct Protocol): This allows a socket application to use the InfiniBand network without changing the application.
SCSI RDMA Protocol: The Small Computer System Interface (SCSI) makes possible data transfer between computers and peripheral devices. It is commonly pronounced "scuzzy." InfiniBand can enable a SCSI system to use RDMA semantics[4] to connect to storage.
IP over InfiniBand: This makes it possible for an application hosted by InfiniBand to communicate to the outside word using IP based semantics.
NFS-RDMA: This is the Network File System over RDMA. The NFS is a widely used file system for use with TCP/IP networks. It allows a computer acting as a client to access files over a network in much the same way it access local files. It was originally developed by SUN Microsystems.
Lustre Support: Lustre is a large scale (massive) file system for use in large cluster computers. Lustre can support tens of thousands of computers, petabytes of storage, and hundreds of gigabytes/second of input/output throughput. It is used in both supercomputers and data centers (http://www.wikipedia.org). It is available under GNU General Purpose License (GPL). The name "Lustre" comes from the words Linux and Cluster.

3.5 Other RDMA Implementations

Two other networking implementations of Remote Direct Memory Access (RDMA) are RDMA over Converged Ethernet (RoCE) and the internet Wide Area RDMA Protocol (iWARP) (http://www.wikipedia.org).

3.5.1 RoCE

As the name implies, RDMA over Converged Ethernet enables RDMA over an Ethernet network. One option, RoCE v1, has the Ethernet protocol acting as a link layer protocol. Two stations using it have to be on the same Ethernet domain. The other option, RoCE v2 can be called routable RoCE or RRoCE. It operates above UDP/IP and can be routed.

One aim of RoCE was to offer an Ethernet based alternative to Infiniband. Infiniband historically offers higher throughput and lower latency than Ethernet at particular points in time, though.

[4] "Semantics" is the meaning of computer instructions—"syntax" is their format.

3.5.2 iWARP

The internet Wide Area RDMA Protocol was specified by the Internet Engineering Task Force (IETF) in a number of documents released since 2007. It allows RDMA over Internet Protocol (IP) networks. It is "layered" (http://www.wikipedia.org) on IETF protocols such as TCP and STCP. The actual zero copy transmission is made possible by the user of the Direct Data Placement Protocol (DDP). However, the transmission is made by protocols such as TCP and STCP, not by DDP.

3.6 Conclusion

InfiniBand competes with Ethernet to provide high performance interconnects for HPC systems. Traditionally HPC was the primary market for InfiniBand. However, it is finding a new role in cloud computing environments including those used for financial services where low latency and high bandwidth are important considerations.

Chapter 4
Wireless Networks

4.1 Introduction

Wireless technology has unique capabilities to service mobile nodes and establish network infrastructure without wiring. Wireless technology has received an increasing amount of R&D attention in recent years. In this chapter, the popular 802.11 WiFi, 802.15 Bluetooth, and LTE standards are discussed.

4.2 802.11 WiFi

The IEEE 802.11 standards [49, 62, 72] have a history that goes back a number of years to the first standard release in 1997. The original standard was 802.11 (circa 1997). However, it was not that big a marketing success because of a relatively low data rate and relatively high cost. Future standardized products (such as 802.11b, 802.11a, 802.11g, and 802.11n) were more capable and much more successful. We will start by discussing the original 802.11 standard. All 802.11 versions are meant to be wireless local area networks generally with ranges of several hundred feet.

4.2.1 The Original 802.11 Standard

The original 802.11 standard can operate in two modes (see Fig. 4.1). In one mode, 802.11 capable stations connect to access points that are wired into a backbone. The other mode, ad hoc mode, allows one 802.11 capable station to connect directly to another without using an access point.

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_4

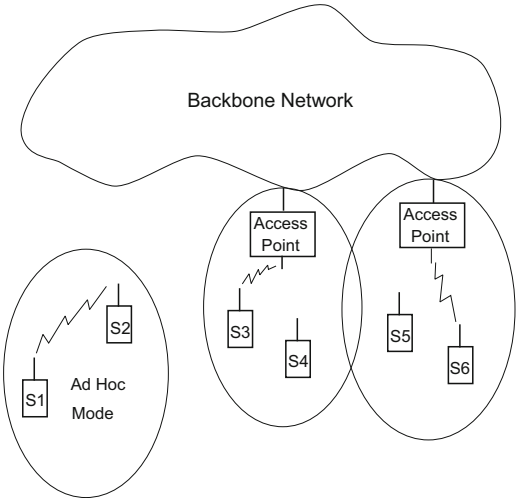


Fig. 4.1 Modes of operation for 802.11 protocol

The 802.11 standard uses part of the ISM (industrial, scientific, and medical) band. The ISM band allows unlicensed use, unlike much other spectrum. It has been popular for garage door openers, cordless telephones, and other consumer electronic devices. The ISM band includes 902–928 MHz, 2.400–2.4835 GHz, and 5.725–5.850 GHz. The original 802.11 standard used the 2.400–2.4835 GHz band.

In fact, infrared wireless local area networks have also been built but are not used today on a large scale. They use pulse position modulation (PPM).

The 802.11 standard can use either direct sequence or frequency hopping spread spectrum. Frequency hopping systems hop between 79 frequencies in the USA and Europe and 23 frequencies in Japan. Direct sequence achieves data rates of 2 Mbps while frequency hopping can send data at 1 or 2 Mbps in the original 802.11 standard.

Because of the spatial expanse of wireless networks, the type of collision detection used in Ethernet would not work. Consider two stations, station 1 and station 2, that are not in range of each other. However, both are in range of station 3. Since the first two stations are not in range of each other, they could both transmit to station 3 simultaneously upon detecting an idle channel in their local geographic region. When both transmissions reach station 3, a collision results (i.e., overlapped garbled signals). This situation is called the hidden node problem [149].

To avoid this problem, instead of using CSMA/CD, 802.11 uses CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). To see how this works, consider only station 1 and station 3 and suppose station 1 wants to transmit to station 3. Station 1 issues an RTS (request to send) message which reaches station 3 which includes the source and destination addresses, the data type, and other information. Station 3, upon receiving the RTS and wishing to receive a communication from station 1, issues a CTS (clear to send) message signaling station 1 to transmit. In the context of the previous example, station 2 would hear

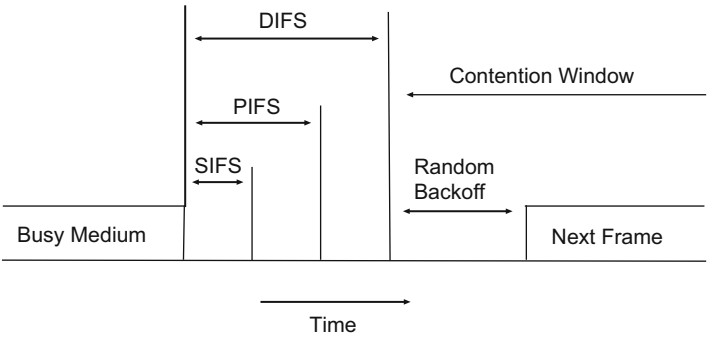


Fig. 4.2 Timing of 802.11 between two transmissions

the CTS and not transmit to station 3 while station 1 is transmitting. Note RTSs may still collide but this would be handled by rescheduled transmissions.

The 802.11 protocol also supports asynchronous and time critical traffic as well as power management to prolong battery life.

Figure 4.2 shows the timing of events in an 802.11 channel. Here after the medium becomes idle a series of delays called spaces are used to set up a priority system between acknowledgements, time critical traffic, and asynchronous traffic. An interframe space is an IFS.

Now, *after* the SIFS (short interframe space) acknowledgements can be transmitted. *After* the PIFS (point coordination interframe space) time critical traffic can be transmitted. Finally, *after* the DIFS (distributed coordination interface space) asynchronous or data traffic can be sent. Thus, acknowledgements have the highest priority, time critical traffic has the next highest priority, and asynchronous traffic has the lowest priority.

The original 802.11 standard had an optional encryption protocol called WEP (wired equivalent privacy). A European potential competitor to 802.11 was HiperLAN1 (and HiperLAN2). However, 802.11 was introduced first and HiperLAN was not a success. The physical layer of 802.11a is very close to that of HiperLAN2, though. Finally, note that wireless protocols are more complex than wired protocols, for local area network and other environments.

4.2.2 Foundational 802.11 Versions

Since the original 802.11, a number of improved versions were developed and have become available. The original 802.11 version itself did not sell well as the price and performance was not that appealing. The four general foundational variations are 802.11b, 802.11a, 802.11g, and 802.11n. Each is now briefly discussed. See Kapp [62], Vaughan-Nichols [155], Poole [116], and Kuran [70] for discussions. More US households use WiFi for a home LAN rather than Ethernet.

Table 4.1 UNI bands

Name	Band
UNI-1	5.2 GHz
UNI-2	5.7 GHz
UNI-3	5.8 GHz

802.11b: This 1999 version first made WiFi popular. It operates at a maximum of 11 Mbps at a range of 100–150 feet and 2 Mbps at 250–300 feet. Data rate decreases with distance to 1 Mbps and then goes to zero. If Wired Equivalent Privacy is used with encryption, the actual useful data rate drops by 50%.

The 802.11b signal is in the 2.4 GHz band. It can operate either using direct sequence spread spectrum, frequency hopping, or infrared. Direct sequence is very popular and infrared is mostly not in use.

802.11a: In spite of the name, 802.11a was developed after 802.11b (Table 4.1). It operates at 54 Mbps in the UNI (Unlicensed Infrastructure Band):

There is some disagreement in the technical literature as to whether 802.11b or 802.11a has the larger range.

802.11g: Sometimes 802.11g is known as 802.11b extended. Initial versions were at 22 Mbps, later versions were at 54 Mbps. Using 802.11g and methods to increase throughput can offer data rates of 100–125 Mbps [70].

802.11n: A committee was formed to create a higher rate version of 802.11 [116] in 2004. Two years later there was early industry agreement on the features of 802.11n. The most significant feature of 802.11n is the use of MIMO (multiple input multiple output) antenna technology. Normally when a radio signal is transmitted from a transmitter to a receiver, the receiver may receive multiple time shifted versions of the original signal. This is because the signal may bounce of different objects and be reflected to the receiver along paths of different lengths. This is called multipath interference. While normally a nuisance, MIMO uses multiple antennas to achieve the transmissions of multiple spatial channels in parallel. This boosts throughput. The throughput of 802.11n can theoretically be as high as 600 Mbps but this is with 40 MHz channels (20 and 40 MHz channels can be used) and with four spatially parallel streams. In practice data rates are often lower than 600 Mbps. The maximum range of 802.11n is 50 m. The modulation scheme used is OFDM (orthogonal frequency division multiplexing).

Note access points can include options for more than one 802.11 version.

A number of specialized 802.11 standards have also been developed (Table 4.2). Some are:

One of these that it would be good to discuss is:

802.11s: Traditionally devices such as laptops are wirelessly connected to access points. These access points are hard wired into the external Internet. The standard 802.11s provides support to allow wireless units such as access points to form wireless mesh networks [48, 53] (http://www.wikipedia.org). This is support. The actual communication is still handled by 802.11 a, b, g, or n. A working group

Table 4.2 Some other 802.11 protocols

Name	Description
802.11e	With quality of service
802.11h	Standard for power use and radiated power
802.11i	Uses WEP2 or AES for improved encryption
802.11p	Vehicular environments
802.11s	Mesh technology support
802.11u	Internetworking support for external networks
802.11x	Light weight version of EAP (Extended Authentication Protocol)

was set up by IEEE to create this standard in 2004. There have been thousands of comments and several delays and a stable version was created around 2010 and approved in 2011 [48].

In 802.11s the basic unit is a "mesh station" which is networked with other mesh stations statically or in an ad hoc fashion. A collection of mesh stations and the networked links between them is a mesh basic service set (MBSS). A mesh station can be a gateway or bridge to an external network in which case it can be referred to as a mesh portal.

Techniques are used in 802.11s that hide the multi-hop nature of the mesh from the upper layers of the protocol stack. It seeks to be functionally equivalent to a broadcast Ethernet.

There is a mandatory path (route) selection algorithm and a mandatory link metric for inter-operability reasons but other algorithms/metrics can be utilized. Unicast, multicast, and broadcast transmission of frames is supported in 802.11s. Mesh stations can detect each other using passive scanning (noticing beacon frames) or active scanning (sending probe frames). Stations may user single transceivers (on one frequency) or multi-transceivers (on different frequencies). Finally most 802.11s meshes will use the 802.11u standard for authentication of users.

A word is in order on 802.11 security. A user requires some sophistication to prevent snooping by others. For instance, security features on shipped products are often disabled by default. Williams [162] reports that many corporate users are not using or misusing WEP. These are media articles on people driving by in vans tapping into private networks. The 2001 article by Williams describes a series of security weaknesses in 802.11b. Some in the wireless LAN industry feel if one uses recommended security practices along with 802.11 features, security is acceptable.

4.2.3 More Recent 802.11 Versions

WiFi has become increasingly ubiquitous. The ever increasing demand for higher data rates has inspired a large number of efforts to extend the capabilities, particularly throughput, of 802.11. A number of these efforts are outlined below [19, 146]. See these excellent references for further discussion.

802.11ac: The 802.11 protocol 802.11n has a 600 Mbps theoretical throughput (which is often lower). The "amendment" IEEE 802.11ac is designed for a throughput close to 1 Gbps. It supports bigger channel widths than 802.11n. In 802.11n the maximum bandwidth is 40 MHz. This is created by "bonding" (or aggregating) two 20 MHz channels. For 802.11ac the channel bandwidth can be 20, 40, 80 (mandatory), and 160 MHz (optional) by bonding a group of adjacent 20 MHz channels. There is also an option to bond two non-adjacent 80 MHz channels which is known as 80+80 MHz [19, 146].

Another feature of 802.11ac is the use of MU-MIMO (multiple user—multiple input multiple output) antenna systems. In single user SU-MIMO (as in 802.11n) data for a single receiver is split into multiple independent spatial streams and transmitted concurrently using multiple antennas. Under MU-MIMO an access point can transmit multiple spatial streams to multiple users concurrently. While devices such as phones or tablets usually have only one or two antennas due to space considerations, access points often have three or four antennas. The number of spatial streams is limited by the total number of antennas in the clients, not "per client" [146]. Only downlink concurrency (from access point to device) is allowed in 802.11ac. Under 802.11ac an access point can have a maximum of eight antennas and transmit four or less spatial streams to two distinct users concurrently or two or less spatial streams to four distinct users concurrently [19].

Another new feature of IEEE 802.11ac is the use of 256-QAM (quadrature amplitude modulation). Briefly, in sending digital signals over a channel, different waveforms (different in amplitude and phase) are used by modulation schemes to represent the binary information. While more throughput in terms of bits per second can be sent, the more distinct waveforms are used, this means the waveforms closely resemble each other leading, in the presence of noise and distortion, to receiver errors (i.e., the receiver mis-identifying the waveform and thus the bits transmitted). While 802.11n uses 64-QAM, 802.11ac uses 256-QAM (with 256 distinct waveforms). With a 160 MHz bandwidth, a data rate of 866.7 Mbps can be obtained with the use of 256-QAM and a single spatial stream. With a maximum of eight distinct spatial streams a throughput of 6.98 Gbps can be achieved [146].

802.11ad: As time has gone on frequencies have become more crowded, leading to a migration of technology to higher frequencies. Thus IEEE 802.11ad (also known as "WiGi") uses the 60 GHz unlicensed frequency region for operation. This is the millimeter-wave (mmWave) band. WiGi utilizes frequencies between 57 and 66 GHz. Planned frequency allocations for 802.11ad for different countries are shown in Table 4.3 [146].

On the 60 GHz bands the propagation loss and signal attenuation is worse than when using the 2.4 and 5 GHz bands for other 802.11 variants. Potential communication distance (range) is thus shorter. On the plus side the bandwidth is larger in the 60 GHz band compared to lower frequency bands. Moreover many antennas can be used in a small area to form a high directional beam to compensate for signal attenuation loss. It should be noted that 802.11ad can reflect off but not penetrate walls, leading to the use of lower frequency 802.11 protocols to do

Table 4.3 802.11ad bands

Country	Band
Korea	57–64 GHz
US, Canada	57.05–64 GHz
EU	57–66 GHz
Japan	59–66 GHz
China	59–64 GHz
Australia	59.4–62.9 GHz

so if a client wanders off from an access point (http://www.wikipedia.org). It is expected that 802.11ad will be used for high definition video and high rate data synchronization.

In terms of the physical layer, 802.11ad supports a single carrier data rate up to 4620 Mbps (using $\pi/2$-16-QAM, 3/4 rate coding, and a data symbol rate of 1540 Msym/s). An OFDM (orthogonal frequency division multiplexing) version is also supported with data rates up to 6756.75 Mbps (using 64-QAM, 13/16 rate coding, 336 data sub-carriers, and an OFDM symbol rate of 4125 Ksym/s) [146]. Both versions are implemented in 2.16 GHz wide channels.

Note that the MAC of the 802.11ad protocol defines time division multiple access on top of existing contention oriented CSMA/CA (carrier sense multiple access with collision avoidance) to achieve quality of service (QoS).

802.11ax: A great deal of effort has gone into boosting the peak throughput of 802.11 protocols operating in a single BSS (basic service set) mode. However, as more access points are deployed in small areas one has an environment of overlapping basic service sets (OBSS). This leads to inter-BSS collisions and contention so that the capacity and coverage goals are not met.

Ongoing work on IEEE 802.11ax (to be released in 2019) seeks to enhance the area throughput and average personal throughput in indoor and outdoor crowded environments by improving the physical and MAC layers.

Challenges in developing IEEE 802.11ax include [19]:

- Make possible at least a four times improvement in capacity with respect to 802.11ac.
- Support for dense networks involving multiple overlapping wireless LANs and stations. There is a need for spatial reuse of transmission resources.
- Minimize the use of management and control packets and make more efficient the packet structure, channel access policies, and retransmission techniques.
- Backwards compatibility. Make mandatory the transmission of the legacy physical layer preamble and keep EDCA (Enhanced Distributed Channel Access) as the channel access policy.
- Minimize energy use.
- Further develop MU-MIMO and OFDM capabilities, in the uplink (station to access point) as well as the downlink.

Next generation wireless LANs will probably implement new functionality. This includes quick, efficient, and robust handoffs between access points located in the same administrative area (domain), device to device (D2D) communication, and coordination of multiple access point networks [19]. The amendment **802.11ai** known as Fast Initial Link Set-up (FILS) addresses the first issue and should be finished in 2016. The goal of 802.11ai is for a handoff to be done in under 100 ms. Communication that is D2D eliminates an access point relay and thus reduces the number of packet transmissions. Finally, a novelty in the management of multiple access points is the virtualization of network functionality. This should help in dense networks and improve the "user experience" [19]. A discussion of possible directions for 8021.ax appears in this reference.

802.11aa: The IEEE 802.11aa amendment, released in 2012, was created to improve the performance of real-time multimedia content transmission. This is done through new features and mechanisms. Older 802.11 standards are lacking in efficient/robust delivery of audio and video streaming.

Several shortcomings of the older 802.11 standards are targeted by 802.11aa. These shortcomings include [19]:

- No support that is efficient and reliable for multicast and group communication.
- Traffic prioritization cannot be applied to distinct multimedia streams or distinct categories of frames in the same stream.
- No techniques for cooperative resource sharing among nearby access points.
- No graceful degradation of audio and video streaming quality.
- Non-consistency/inter-operability with the older 802.1 protocol for audio video bridging (AVB).

Solutions and challenges for these problems are discussed in [19].

802.11af: With the transition from analog to digital television, spectrum in the very high frequency band (VHF) and the ultra high frequency band (UHF) has become available to unlicensed users as long as they don't interfere with licensed users. This spectrum is known as *TV white space* (TVWS). This white space exists in contiguous 470–790 MHz in Europe, contiguous 470–710 MHz in Japan, and non-contiguous 54–72, 76–88, 174-216, and 470–698 MHz in the USA and Korea [146]. This opportunity has led to IEEE 902.11af, published in 2014 [19].

Notice that all these white space frequencies are below 1 GHz. Propagation characteristics are superior at these frequencies compared to the 2.4 GHz, 5 GHz, and higher bands usually used and considered for 802.11. Below 1 GHz signals travel longer distances with less path loss and an improved ability to go through walls. Access points need not be densely deployed. Applications include Internet access in sparsely populated areas, smart grid applications, sensor networks, metering, the Internet of things, and TVWS traffic off-loading in indoor environments [19].

Geolocation databases (GDBs) are used to store location based information on available spectrum and usage schedules. Location aware access points under 802.11af access the Internet to determine where they can operate, what system they can use, and when they can operate so as not to interfere with licensed devices.

Among specifications that IEEE 802.11af meets is narrow channel bandwidth (6–8 MHz depending on the country). It also operates with non-contiguous time-varying available channels due to usage by digital TV users. The IEEE 802.11af amendment uses features of 802.11ac such as MU-MIMO since its physical layer is built on 40 MHz 802.11ac. Both contiguous and non-contiguous channel bonding of up to four channels is possible. Spectrum guard bands provide protection for TV users on adjacent channels [146].

802.11ah: Another below 1 GHz 802.11 protocol, IEEE 802.11ah, uses unlicensed spectrum to provide long distance, lower data rate, connectivity for applications such as sensor networks, the smart grid, and the Internet of things. Another reason for considering 802.11ah is a fear that 802.11af may have problems finding bandwidth in urban areas with many (licensed) TV stations. Table 4.4 shows frequency bands for 802.11ah [146].

The 802.11ah protocol is suited for large scale, low data use sensor networks. Since many of the devices may be battery powered, minimizing energy usage is important. Obtaining better spectral efficiency is also important because of the limited bandwidth below 1 GHz [146].

Requirements of 802.11ah include [19]:

- A single access point can support 8192 (2^{13})stations.
- Efficient power saving techniques.
- Operation in unlicensed bands below 1 GHz.
- Minimum data rate of 100 kbps.
- Outdoor range of up to 1 km.

Channel widths of 1 and 2 MHz have been specified. Some countries support channel widths of 4, 8, and 16 MHz. New physical and MAC layers have also been defined for 802.11ah. The 802.11ah physical layer is in some sense a below 1 GHz implementation of 802.11ac. The 802.11ah MAC protocol utilizes most of 802.11's main characteristics and enhances its power savings mechanisms.

Table 4.4 Freq. bands of 802.11ah

Country	Band
US	902–928 MHz
Korea	917.5–923.5 MHz
EU	863–868 MHz
China	755–787 MHz
Japan	916.5–927.5 MHz
Singapore	866–869, 920–925 MHz

4.3 802.15 Bluetooth

The original goal of Bluetooth technology, standardized as IEEE 802.15, is to provide an inexpensive, low power chip that can fit into any electronic device and use ad hoc radio networking to make the device part of a network. For instance, if your PC, printer, monitor, and speakers were Bluetooth enabled, most of the rat's nest of wiring under a desk top would be eliminated. Bluetooth chips could also be placed in PDAs, headsets, etc. Bluetooth has been deployed in a large number of personal area networks (PAN) and personal devices.

Work on Bluetooth started in 1997. Five initial corporate supporters (Erickson, Nokia, IBM, Toshiba, and Intel) grew to a more than a thousand adopters by 2000. The name Bluetooth comes from the Viking King of Denmark, Harald Blatand, who unified Norway and Denmark in the tenth century [51].

4.3.1 Technically Speaking

Bluetooth had a number of design goals. As related in Haartsen [51], among these are:

- System should function globally.
- Ad hoc networking.
- Support for data and voice.
- Inexpensive, miniature, and low power radio transceiver.

The original Bluetooth operates in the 2.4 GHz ISM band (see the previous section's discussion of the ISM band). It uses frequency hopping spread spectrum (79 hopping channels, 1600 hops/s). Time is divided into 625 μs slots with one packet fitting in one slot. The data rate is 1 Mbps. The range is 10 m, making Bluetooth a personal area network (PAN).

Two types of connections are possible with Bluetooth. First, SCO links (Synchronous Connection Oriented) are symmetrical, point to point, circuit switched voice connections. Secondly, ACL links (Asynchronous Connectionless) are asymmetrical or symmetrical, point to multipoint, packet switched connections for data.

A number of features of Bluetooth are designed to make possible good interference immunity. One is the use of high rate frequency hopping with short packets. There is an option to use error correction and a fast acting automatic repeat request scheme using error detection. Finally, voice encoding that is not susceptible to bit errors is used.

4.3.2 Ad Hoc Networking

Two or more Bluetooth nodes form a "piconet" in sharing a frequency hopping channel. One node will become a "master" node to supervise networking. The other nodes are "slaves". Not only may roles change but roles are lost when a connection is finished.

All SCO and ACL traffic is scheduled by the master node [51]. The master node allocates capacity for SCO links by reserving slots. Polling is used by ACL links. That is, the master node prompts each slave node in turn to see if it has data to transmit (see Schwartz [132] for a detailed discussion of polling). Slave node clocks are synchronized to the master node's clock.

There is a maximum of eight active nodes on a single piconet (others may be parked in an inactive state). As the number of nodes increases, throughput (i.e., useful information flow) decreases. To mitigate this problem, several piconets with independent but overlapping coverage can form a "scatternet." Each piconet that is part of a scatternet uses a separate pseudo-random frequency hopping sequence. The scatternet approach results in a very small decrease in throughput. Note that a node can be on several piconets in a scatternet. Moreover, it may be a master node on some piconets and a slave node on other piconets.

4.3.3 Versions of Bluetooth

There are five versions of Bluetooth as illustrated in Table 4.5.

The indicated 25 Mbps data rate in the table is not transmitted over Bluetooth itself. Rather it is transmitted over a parallel 802.11 link whose operation is negotiated using Bluetooth. Version 4.0 includes protocols for (a) classic Bluetooth, (b) Bluetooth high speed, and (c) Bluetooth low energy. Version 5.0 was publicized in June 2016 with the first products expected in the late 2016 or the early 2017. It seeks to increase the range by a factor of four, double the speed, and increase throughput by a factor of eight for low energy Bluetooth. Applications are expected to be related to the Internet of Things (http://www.wikipedia.org).

There was some effort that was not realized to make an 802.15 variant ultra wideband (UWB) standard for Bluetooth [156]. Ultra wideband technology is a

Table 4.5 Bluetooth versions [Wikipedia]

Version	Data rate (Mbps)	Max application throughput (Mbps)
Version 1.2	1	0.7
Version 2.0+	3	2.1
Version 3.0+	25	2.1
Version 4.0	25	2.1
Version 5.0	50	

short range radio technology that spreads the communication spectrum over an unusually wideband of spectrum with very narrow pulses. It can have low energy requirements.

4.3.4 802.15.4, ZigBee, and 802.15.4e

In actuality, the original Bluetooth faced some problems in gaining acceptance. The rapid growth of 802.11 technology and its pricing had not given Bluetooth a price advantage on certain applications [171]. Also, Bluetooth is more complex than its original design goal as an attempt was made to have it serve more applications and supply quality of service. There was also some question on the scalability of scatternets. However, Bluetooth eventually did gain acceptance.

In this section low data rate extensions of Bluetooth are discussed.

A key component of the Internet of Things (IOT) is wireless sensor and actuator networks (WSANs). This is because WSANs are the way in which a computer system can interface with the physical world. Sensors bring data on the environment and component status to the IOT computational system while actuators allow the computational systems to influence/direct the IOT components. Applications for WSANs, many of them current, are in areas such as health care, the smart grid, monitoring involving safety, inventory tracking, factory and warehouse automation, entertainment, toys, and games. Requirements for WSANs include reliability, timeliness, scalability, and energy effects [38, 171].

IEEE 802.15.4 is a protocol for low rate wireless networks. It specifies a physical layer and media access control. It is the foundation for the ZigBee, ISA 100.11a, WirelessHART, MiWi, and Thread protocols. These protocols speak to the upper layers not defined in 802.15.4 [38] (http://www.wikipedia.org). Some of these protocols support the creation of WSANs. Also, the Internet Engineering Task Force (IETF) has designed protocols to integrate smart objects (i.e., sensor/actuator devices) into the Internet including IPv6 over low power WPAN (6LoWPAN) adaptation layer protocol, the Routing Protocol for Low Power and Lossy Networks (RPL), and the Constrained Application Protocol (CAP) [38].

Channels and frequency bands for 802.15.4 are shown in Table 4.6 (http://www.wikipedia.org).

Table 4.6 IEEE 802.15.4 channels

Area	Frequency bands	Number of channels
Europe	868.0–868.6 MHz	1
North America	902.0–928.0 MHz	30 (as of 2006)
World Wide	2400–2483.5 MHz	16

4.3.4.1 ZigBee

ZigBee is built on an IEEE 802.15.4 foundation. While it is widely used in applications involving metering it is not suitable for industrial applications with real-time requirements. Two industry standards also built on 802.15.4 are WirelesHART and ISA100 [168]. They emphasize real-time and coexistence functionality.

ZigBee, like Bluetooth, has a range of 10 m. Communication can take place from a device to a coordinator, a coordinator to a device, or between stations of the same type (i.e., peer to peer) in a multi-hop mode of operation. An 802.15.4 network can have up to 64,000 devices in terms of address space. ZigBee topology includes a one hop star or the use of multi-hopping for connectivity beyond 10 m.

In beacon-enabled mode, the coordinator periodically broadcasts "beacons" to synchronize the devices it is connected to and for other functions. In non-beacon-enabled mode, beacons are not broadcast periodically by the coordinator. Rather, if a device requests beacons, the coordinator will transmit a beacon directly to the device [171]. A loss of beacons can be used to detect link or node failures.

It is critical for certain applications to minimize ZigBee coordinator and device energy usage. Some of these applications will be battery powered where batteries will not be (practically or economically) replaceable.

The majority of power savings functions in 802.15.4 involve beacon-enabled mode. In direct data transmissions between coordinators and devices, the transceivers are only on 1/64 of the duration of a packetized super-frame (i.e., collection of slots). A small CSMA/CD backoff duration and brief warm-up times for transceivers are also used to minimize power usage in 802.15.4.

Three security levels are available in 802.15.4. The lowest level is None Security mode which is suitable if the upper layers provide security or security is not important. An access control list is used in the second level of security to allow only authorized devices to access data. The Advanced Encryption Standard (AES) is used in the highest, third security level.

4.3.4.2 IEEE 802.15.4e

As has been just mentioned, many protocols have been developed to provide networking support to WSAN applications. The basis of many of these protocols is IEEE 802.15.4. It has been recognized though that 802.15.4 has a number of limitations and constraints.

802.15.4 Limits: IEEE 802.15.4e addresses a number of the deficiencies of IEEE 802.15.4. These deficiencies include [38]:

- Unbounded Delay: There is no bound on maximum delay for packets as 802.15.4 uses CSMA/CA (see Sect. 4.2.1). This is true in both beacon-enabled (BE) mode and non-beacon-enabled (NBE) mode.

- Communication Efficiency: Due to the CSMA/CA algorithm used there is a low packet delivery ratio even for small networks in BE mode and in NBE mode when a large number of nodes begin transmitting concurrently (say in response to an event).
- Sensitivity to Fading/Interference: Whereas some networks (such as Bluetooth, ISA 100.11a, and Wireless HART) use frequency hopping to reduce the effects of fading and interference, 802.15.4 has no frequency hopping mechanism and uses a single channel.
- Relayed Communication: Establishing multi-hop networks in 802.15.4 BE mode involves complex synchronization and beacon scheduling (not included in the standard). As an alternative relay nodes in 802.15.4 multi-hop topologies have their radio activated all of the time, leading to a large energy usage.

The 802.15.4e Solution: To provide an improved protocol that addressed these limitations in 2008 a working group was created to improve the 802.15.4 MAC protocol in the context of embedded applications (i.e., industrial applications). The resulting amendment IEEE 802.1.5.4e was released in 2012. To some extent it is based on current standards for industrial applications such as ISA 100.11a and WirelessHART. It uses concepts from these earlier protocols such as frequency hopping, multi-channel communications, slotted access, and dedicated slots. The new protocol introduces MAC improvements that are not connected to specific application areas (called general functional enhancements). The new protocol also introduces new MAC Protocols intended to support certain application areas (called MAC behavior modes). These are now outlined.

General Functional Enhancements: The general functional enhancements introduced in 802.15.4e are [38]:

- Low Energy (LE): This enhancement is for applications that can "trade" latency for energy efficiency. A node can have a low duty cycle (i.e., percentage of on time), say 1% or below while at the same time looking to upper layers like the node is always on. This is relevant to the Internet of Things as Internet protocols are designed with the assumption that hosts are always on.
- Information Elements (IE): A mechanism that can be extended to allow a MAC layer exchange of information.
- Enhanced Beacons (EB): An extension allowing greater flexibility to 802.15.4 beacon frames. Application specific beacon frames can be developed by using relevant information elements.
- Multipurpose Frame: Using IEs as a foundation, this enhancement makes available a versatile frame format that can support a number of MAC operations.
- Performance Metric: Gives feedback to the networking/upper layers on channel quality. This allows informed decisions to be made.
- Fast Association (FastA): To minimize energy usage, the 802.15.4 association procedure causes a delay. But in some situations (i.e., those that are time critical) delay is more important than minimizing energy usage. The FastA enhancement makes possible association with reduced latency.

MAC Behavior Modes: There are five new MAC behavior modes in IEEE 802.15.4e [38]:

- Time Slotted Channel Hopping (TSCH): Using a TDMA approach, it supports multi-channel and multi-hop communications. Intended applications include process control and industrial automation.
- Deterministic and Synchronous Multi-Channel Extension (DSME): Puts together time division and contention based medium access. It supports two distinct channel diversity modes. Intended for applications with tight requirements of reliability and timeliness arising in both commercial and industrial applications. It is designed especially for mesh and multi-hop networks.
- Low Latency Deterministic Network (LLDN): This is for single channel, single hop networks used in factory automation applications demanding very low delay.
- Asynchronous Multi-Channel Adaptation (AMCA): Intended for large deployment applications (e.g., smart grids, process control networks, and infrastructure monitoring). In these network applications using a single channel may not make it possible to connect all nodes in the same personal area network (PAN), the variability in channel quality is large, and there may be link asymmetries between adjacent nodes (i.e., a link is uni-directional). The AMCA enhancement is based on asynchronous multi-channel adaptation and can only be used in non-beacon-enabled networks.
- Radio Frequency Identification Blink (BLINK): Enables a node to send its ID to other nodes without a need for prior association and also without acknowledgements. It uses the ALOHA protocol [122]. Intended applications include location and tracking and also item or person identification. Extended discussions of TSCH, DSME, and LLDN appear in the excellent survey [38].

4.3.5 Wireless Body Area Networks and 802.15.6

4.3.5.1 Introduction

Wireless body area networks (WBAN) are networks that can collect and transmit data from different parts of the human body, possibly to remote sites [71, 73]. Applications include health care, finding lost items, data file transfer, sports, gaming and entertainment (sensors can collect data on body part movement and status), social networking (for instance, exchanging electronic business cards by simply shaking hands), and military uses [27].

Health care may be the biggest application of wireless body area networks. An obvious application is to collect vital patient data and forward it to a remote site for processing and diagnosis. This could be used for myocardial infarctions, and also, for instance, for diseases involving cancer, asthma, the neural system, and the gastrointestinal tract [71].

Several protocols could be used for wireless body area networks. These include generic Bluetooth (802.15), ZigBee, Bluetooth Low Energy, and 802.15.6. Battery life and sharing the radio environment with other wireless devices/networks are crucially important. Radio propagation is influenced by the human body which leads to a unique radio channel which has to be taken into account.

A good number of WBAN implementations operate in the ISM (industrial, scientific, and medical) band centered about 2.45 GHz. The ISM band is heavily used by other wireless networks such as WiFi 802.11. There are other bands dedicated to medical devices. These include [27]:

- Wireless Medical Telemetry System (WMTS).
- Medical Implant Communication (MICS).
- Medical Body Area Networks (MBAN).

The first two were allocated for body-worn and implanted medical devices needing simple point to point connectivity. They were meant to allow superior operation in terms of range, bit rate, and reliability compared to magnetic coupling based communication used in early wireless medical devices. In the MICS band a bit rate of 400 Kbps and a range of 2 m are possible (for instance, for pacemakers). In the WMTS band a bit rate of about 1 Mbps can be achieved (as in a swallowable camera pill).

In the USA in 2012 the FCC (Federal Communications Commission) allocated 40 MHz of spectrum between 2.36 and 2.4 GHz (on a secondary basis) for a new Medical Body Area Network (MBAN) services requiring licenses. This should provide an alternative to the crowded and adjacent ISM band.

There is a wide array of challenges in developing a WBAN. These include [27]:

- Bit Rate/Quality of Service: This can be from 1 Kbps (monitoring temperature) to 10 MBps (video streaming).
- Rates and Topology: A range of no more than a few meters is reasonable. Often a simple star topology is used. The body can block radio communications (i.e., with implants) so multi-hop communications may be needed.
- Security: This is important in medical and military communications. Issues that need to be solved are privacy, confidentiality, authorization, and integrity [27].
- Antenna and Radio Channel: The environment of the human body needs to be taken into account in designing miniaturized antennas. There is a trade-off between antenna size and efficiency.
- Power Consumption: This depends on the specific application(s). Generally WBAN devices are battery powered and battery lifetimes sometimes need to be measured in years (pacemakers need at least 5 years of life) [27].
- Coexistence: With most WBANs operating in the ISM Band, interference from WiFi (IEEE 802.11), Bluetooth (IEEE 802.15.1), and ZigBee and other network technologies is a serious concern. Often WBAN networks require high reliability and the ability to reliably transmit emergency/alarm messages.

- Form Factor: For implantable devices the battery and antenna needs a small and tight case. For implantable devices good radio radiation characteristics and lifetime is also important. Flexibility and stretchability are also important for devices to be worn [27].
- Signal Processing: Power efficient signal processing can minimize the power needed to acquire and process biological signals.
- Health Effects: The potential health effect of WBANs at WBAN frequencies is heating of tissue. General limits on the time-varying electromagnetic fields are specified by the International Commission on Non-Ionizing Radiation Protection (ICNIRP) (see Cavallari [27] for a more detailed discussion).

4.3.5.2 802.15.6

Work on the IEEE 802.15.6 standard is an attempt to standardize networking for wireless body area networks.

As of about 2010, 802.15.6 provides three physical (PHY) layer specifications [71]. The use of each depends on the type of application need:

(1) Narrowband PHY: This physical layer handles radio transceiver activation and de-activation, data transmission and receiving, and clear channel assessment within the current channel

(2) Ultra Wideband PHY: This has a low and high band. There are eleven potential channels of 499.2 MHz bandwidth each. The low band consists of channels 1 through 3 and the high band consists of channels 4 through 11. Channel 2 has a center frequency of 3.9936 GHz and channel 7 has a center frequency of 7.9872 GHz. Channel 2 and channel 7 are "mandatory" channels, other channels do not have to be used. Typical data rates are 0.5–10 Mbps.

Ultra wideband transceivers are relatively simple.

(3) Human Body Communication PHY: There is a band centered at 21 MHz with a bandwidth of 5.25 MHz [27].

A super-frame [1] structure is used by 802.15.6. Beacon periods (of the same length as a super-frame) bound each super-frame. Boundaries of the beacon period are selected by the hub station which is thereby allocating time slots.

Like ZigBee, with 802.16.6 there are security options for (a) unsecured communication (b) authentication only, and (c) authentication and encryption.

4.3.6 Bluetooth Security

Hackers and researchers have discovered several security weaknesses in Bluetooth enabled devices [39]. In fact most Bluetooth attacks are not detected and are more localized than Internet attacks so that they do not get the same amount of attention

[1] A super-frame is a sequential concatenation of frames that repeats periodically.

from the public. It should also be noted that mobile and embedded Bluetooth devices that receive attacks have few or no security features.

Dunning [39] holds that one should use Bluetooth if proper security is in place.

Most of the susceptibility of Bluetooth devices to attacks comes from lax default security settings, a lack of understanding on the part of Bluetooth device owners of security practice and deficient software development. One is safe from most Bluetooth attacks if security settings are correctly configured.

Bluetooth has a few weaknesses that are inherent [39]:

- Wireless data can be (locally) intercepted.
- No third party can verify addresses, names, and classes as on the Internet.
- Many devices can't be patched so any weaknesses remain as long as they are in use.

While Bluetooth threats will probably increase, the key to making them less effective is a better understanding of their potential.

4.4 802.16 WiMax

The original 802.16 protocol allowed connectivity between base stations and fixed computers within buildings. Later work allowed connectivity to mobile users (802.16e in 2005). For some time 802.16 versions such as 802.16m provided a viable framework for 4G cellular communications. It was eventually supplanted by LTE technology. For instance, Sprint shut down its WiMax network circa 2015 [44]. For more details of WiMax see Ahmadi [4], Bacioccola [14], Eklund [40], Papapanagiotou [110], Peters [115], Robertazzi [123] and Tanenbaum [149].

4.5 LTE: Long Term Evolution

4.5.1 Introduction

Long Term Evolution or "LTE" is a process to create an air interface for cell phones by the 3rd Generation Partnership Project (3GPP). However, LTE can also be considered to be a system consisting of architecture and protocols and performance goals.

When cell phones first came into use in the early 1980s these "1st generation" phones were analog based. Second generation phones were digital. Third and fourth generations have increasing data rates allowing new services such as web surfing and video. Some people call LTE a 3.9 technology in that it is almost but not quite 4th generation technology. But the more recent "LTE Advanced" is a 4th generation system.

The initial proposal for LTE came from NTT DoCoMo of Japan. Began first as a study item, the technical requirements of LTE were agreed to in 2005. The LTE standard was frozen in 2008. The first working LTE service for the public was in Stockholm and Oslo in 2009. A number of carriers publicized plans, starting in 2009, to transform their networks into LTE networks. These efforts began in 2010. Finally, LTE Advanced was submitted as a candidate system that is fourth generation to ITU-T (International Telecommunications Union—Standardization Sector). The plan was for this to be finalized by 3GPP in 2011 (http://www.wikipedia.org). LTE technology is specified in release 8 and release 9 standards.

4.5.2 LTE

The goals of LTE include [117, 120] reduced latency, higher data rates for customers, better system capacity and coverage, and a lower cost of operation.

Some important LTE features include [5, 120]:

- A flat IP oriented network architecture using distributed servers.
- In LTE, base stations have transport connectivity to the core network without intervening RAN (radio access network) nodes (e.g., radio network controllers).
- Elegant and efficient radio protocols. State information on channels is accessible to radio protocols peers for efficient operation.
- A physical layer design based on processing in the frequency domain (Fast Fourier Transforms (FFTs) are used). This aids efficient operation and helps to support high data speeds. Bins with a 20 MHz width are used with a 2048 point FFT computation.
- The use of multiple antenna transmission.
- Clever resource management of the radio spectrum to enhance scalability of bandwidth and multi-user diversity. For instance, scheduling can be done in the time-frequency domains.
- Power saving mode is an inherent feature of customer equipment.

In terms of some of the system parameters of the LTE system one has [117] (http://www.wikipedia.org):

- Scalable bandwidth from less than 5 to 20 MHz enabling operation over available spectrum [67].
- A peak download rate of 326.4 Mbps for 4×4 antennas and a peak download rate of 172.8 Mbps for 2×2 antennas, both utilizing 20 MHz of spectrum. A peak upload rate of 86.4 Mbps for each 20 MHz of spectrum (with one antenna).
- At least 200 active users in each 5 MHz cell.
- Latency below 10 ms for short IP packets [67].
- In rural regions cell sizes of 5, 30, and 100 km. For urban regions cell sizes can be as small as 1 km or less.

- There are five terminal classes. These include one class that emphasizes voice and a high end terminal class supporting maximum data rates.
- LTE has excellent support for mobility. High data rates are possible at speeds up to 300–500 km/h. This is a function of the frequency band employed.

4.5.3 LTE Advanced

Following LTE is the fourth generation radio technology standard called LTE Advanced [5, 6] (http://www.wikipedia.org). Fundamentally LTE Advanced is intended to allow for higher data rates and transmission speeds for cellular telephone networks. In fact LTE Advanced is backwards compatible with LTE. It uses the same frequency bands as LTE. However, LTE is not compatible with other third generation systems. As one might expect, LTE Advanced needs smaller latencies than LTE.

LTE Advanced technology has been standardized in a number of releases starting with release 10 which was frozen in 2011. Concepts included in the release include carrier aggregation for use in multiple frequency bands, improved multiple input multiple output (MIMO) technology, relays, and self-organizing networks. Release 10 was followed by enhancements in release 11 and release 12 (frozen circa 2012). Additional enhancements should be in release 13 and potential new access technology in release 14 and release 15. Technology from release 12 and forward is referred to as Beyond 4G (that is B4G [6]).

In 2014 South Korea was the only country implementing LTE Advanced. By early 2015, LTE Advanced was available in 31 countries including the UK, Australia, the USA, France, and Germany. At that time 107 telecommunications providers in 54 countries were investing in LTE Advanced. This is about 30% of the operators that have LTE services [119].

Goals for LTE Advanced are peak data rates of 3 GBps on the downlink and 1.5 Gbps on the uplink and spectral efficiencies of 30 bps/Hz in release 10 [157]. However, actual current bit rates may be much lower and the roll out of LTE Advanced networks and cell phone capabilities is an uneven process. Speed is also affected by one's distance to a cell tower and the number of other users in the network. One report from the UK in December 2015 lists a theoretical LTE Advanced download speed of 300 Mbps but an actual download speed of 42 Mbps (4G-UK 2015).

Release 10 has a number of core technologies that are good to discuss [6]:

Carrier Aggregation: This is the first new functionality to be implemented by carriers [119]. It allows a user to download data in multiple bands concurrently. Data carrying parts can have bandwidths of 1.4, 3, 5, 10, 15, and 20 MHz [153]. One can aggregate up to five such "component carriers," for a maximum download rate in release 10 of 100 Mbps. There are three ways (contiguous versus non-contiguous, intra-band versus inter-band) this can be done. See Fig. 4.3.

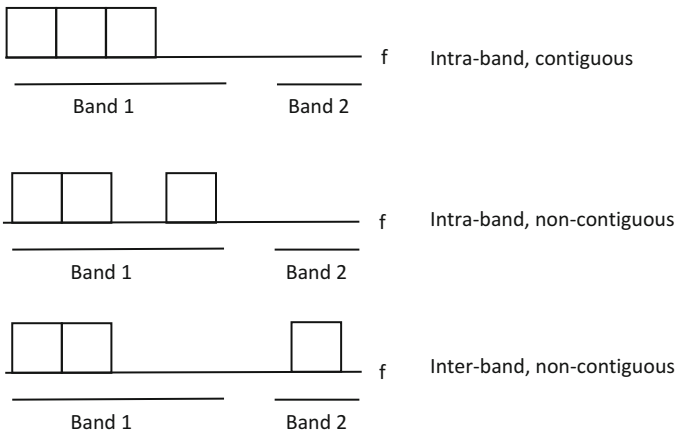


Fig. 4.3 LTE Advanced carrier aggregation modes

Carrier aggregation has benefits besides increasing throughput including inter-cell interference reduction, better handovers, energy savings, and load balancing [6].

MIMO: The use of MIMO (multiple input multiple output) techniques in releases 10, 11, and 12 of LTE Advanced is referred to as "enhanced MIMO," to distinguish it from earlier MIMO implementations as in LTE.

Release 8 and 9 of LTE allow up to 4×4 antenna configurations (4 at the transmitter and 4 at the receiver) on the downlink and 2×2 antenna configurations on the uplink. In the context of LTE there are four basic modes of operation [6]:

- Spatial Diversity: The transmission reliability can be improved by using multiple antennas. That is, different antennas at the transmitter send precoded versions of the same data. The receiver's multiple antennas do different types of signal integration to yield a better signal to noise ratio.
- Spatial Multiplexing: As has been said when discussing 802.11n, wireless signals take multiple paths of different lengths in going from transmitter to receiver because the signal is reflected off surfaces. Concurrent data streams (referred to as "layers") can be sent over these paths using MIMO technology. This parallel transmission of data boosts throughput. It needs precoding at the transmitter to make the data streams orthogonal so they can be easily distinguished from one another. Also, both spatial diversity and spatial multiplexing use cell specific reference signals for demodulation and channel estimation [6].
- Beam Forming: By appropriately adjusting phases of antennas the directivity/sensitivity pattern of a group of antennas can be controlled. This is particularly useful to connect with users at the connectivity region's boundary. For release 12 and higher releases a promising technology is Full Dimension MIMO (FD-MIMO) where a relatively large number of antennas (16, 32, or 64) at the base

station (eNB—"enhanced Node B" in 3GPP language) do accurate 3D beam forming to intended users. Beams can be steered in both azimuth and elevation directions, making it more flexible than two dimensional beam forming.

- Spatial Division Multiple Access (MU-MIMO): System throughput can be boosted if several users can be co-scheduled on the same time and frequency resources. This type of MU-MIMO is supported in LTE release 8 and release 9 based on codebook precoding. Note that this limited approach does not cancel intra-cell interference [6].

Enhancements to LTE MIMO were made in LTE Advanced release 10 and some more refinements appear in release 11. Also, in LTE Advanced the largest number of antennas is 8×8 for the downlink and 4×4 for the uplink. A new feature is a dynamic framework for both SU-MIMO and MU-MIMO switching at the eNB. Here in a fast time scale, the operation mode of different users can be transparently adapted to higher layers [6]. This is useful because of fluctuations in the MIMO channel and traffic amount. Also being investigated is massive MIMO (involving 100 or more antennas).

Cooperative Multipoint Transmission and Reception: This is an important technique to boost coverage and cell boundary throughput and improve the system efficiency. It was discussed in LTE Advanced release 10 and the specification appeared in release 11. The basic concept is that users near the cell boundary can receive downlink transmissions from multiple antennas. Alternately, multiple base stations can coordinate uplink transmissions coming from users. There are several scenarios where such cooperation can take place involving simple schedule coordination up to complex joint signal transmission [6]. One can have cooperation that is intra-site (within the same base station), inter-site (among different base stations) and one can have cooperation in homogeneous or heterogeneous environments.

Relays: A relay node (RN) in LTE Advanced release 10 transfers information between a user ("UE" or user equipment) and base station (eNB). A user transmits to and receives from the relay node using the Uu interface and the relay node transmits to and receives from the base station node using the Un interface. As such, a relay node has both user and base station functionality.

Advantages of using relays include providing coverage in new areas where traditional backhaul (e.g., fiber) is not appropriate, temporary network deployments, boosting cell boundary throughput, improved throughput in regions with poor SNR, or providing coverage in mobile environments such as trains [6].

There are also economic advantages to the use of relays as their cost is less than that of a full base station, one does away with the need for wired backhaul and finally the total power needed to serve the user population should be less than that without relays.

Heterogeneous Networks: To meet growing demand for cellular service, the use of smaller low powered cellular layers has been introduced. Picocells and metrocells are deployed outside while femtocells are deployed inside residential, business, or government buildings. A network with cellular layers with different properties (range, transmission power, backhaul implementation) is called a "heterogeneous network" (i.e., "HetNet").

The concept of HetNets is not new and is not peculiar to LTE Advanced. Three HetNet challenges are cell association (i.e., which cell a user communicates with at a given time), inter-cell interference management (such as using different carrier frequencies for different cell layers), and mobility issues (handover failure rates are higher in HetNets than in homogeneous environments).

Self-Organizing Networks: Such self-organization goes back to LTE release 8. Self-organization techniques have advantages in terms of reducing costs by minimizing manual operation/inputs and speeding the response time to trigger events [6].

Self-organization may involve self-configuration of base stations (eNBs), self-healing of eNBs in response to network problems, and automated discovery of neighbor relations. It may also involve automated reconfiguration of network, system and radio parameters in order to optimize capacity and coverage, interference control, handover parameters, and load balancing. LTE Advanced enhanced self-organization functions began under LTE and new ideas have been studied.

Two other types of communication that involve LTE Advanced are:

Machine to Machine Communication (M2M): This involves machines communicating with each other as part of a network. It is also referred to as Machine type communications. It arises in technological environments such as smart grids and the Internet of Things. It can be implemented by WiFi, ZigBee, Bluetooth, and by mobile cellular networks. The last technology has claimed advantages in terms of ease of installation, reliable connectivity to remote servers, and support for mobility.

A challenge in doing M2M communication for LTE Advanced is that there may be a very large number of devices, each requiring a small amount of bandwidth for data communications (leading to inefficient, large signaling, and data overhead). It may also reduce resources for devices requiring high data rate transmissions. Work on solutions (particularly air interface support for M2M equipment) has been undertaken.

Device to Device Communication: This negates the need for using cellular infrastructure for connectivity between users (though D2D communication may rely on the network for control functions). It has been considered in release 12. This functionality already appears in Bluetooth and WiFi direct but they are not designed to be integrated into the cellular network.

Among applications that would benefit from D2D communication are context aware services where multiple services would be available to users based on user location, public safety when network infrastructure is not available, M2M and the Internet of Things, and finally D2D communication that allows traffic off-loading from the cellular network.

4.5.4 Towards 5G

LTE Advanced is a fourth generation system. The continuing need for providing higher and higher data rates and the push of new and novel technology suggests we are heading into a Fifth Generation (5G) of cellular systems [54, 75, 107].

Some expect a dual track approach over the next few years [75]. Under the "evolution" track the evolution in LTE Advanced will proceed in releases 13, 14 and further. This will be done in a backward compatible fashion with system performance being improved in bands below 6 GHz. At least part of the 5G requirements will thus be met by evolution in LTE Advanced.

Under the second radio access technology (RAT) track there is no requirement for backwards compatibility and revolutionary new technologies can be integrated to yield the best performance. A new RAT 5G system will satisfy all 5G specifications and replace 4G systems in the future. The new RAT technology should be able to support bands both below and above 6 GHz in the years to come [75].

Releases 13 and 14 serve as a "bridge" [75] for 4G to 5G technology. Among release 13 and release 14 features are:

Full Dimension MIMO (FD-MIMO): As mentioned, the basic concept is to use a 2D antenna array panel to form narrow radio beams both horizontally and vertically. Release 13 boosts the number of transmit antennas from 8 to 16 at the base station (eNB). A number of FD-MIMO enhancements were included in release 13. Note that FD-MIMO can also be deployed in indoor, micro, and picocells. Higher frequencies also result in smaller antenna systems as antenna spacing is inversely proportional to carrier frequency [75].
Release 14 will aim for up to 32 transmitting antennas. There are technical challenges in doing this [54].

Licensed Assisted Access (LAA): LTE Advanced systems operate in licensed bandwidth. However, the bandwidth is limited and expensive. Therefore the additional use of unlicensed bandwidth to supplement the use of licensed bandwidth is of much interest [75]. There are challenges in doing this such as coexistence with unlicensed networks such as WiFi.
In release 13 downlink transmission only using unlicensed bandwidth was specified and some uplink functionality was created for uplink use in future releases such as release 14. Licensed and unlicensed bands are used jointly in release 13 via carrier aggregation. An important part of LAA is listen before talk (LBT) functionality that checks that a band is clear (idle) before its use [75].
Release 14 will include aggregation on the uplink. A number of protocol enhancements will be added in release 14 to increase LAA uplink efficiency.

Carrier Aggregation Enhancements: In release 10 the carrier aggregation of up to five component carriers with common FDD and TDD duplexing was specified for a common bandwidth of 100 MHz. The joint use of MIMO and carrier aggregation results in a peak downlink rate of 3 Gbps and a peak uplink rate of 1.5 Gbps. Release 12 extended carrier aggregation to FDD and TDD carrier aggregation but also with at most five carrier components.

A work item was approved for release 13 for carrier aggregation for up to 32 component carriers [75].

Machine Type Communications (MTC): A cost decrease of about 50% for MTC user equipment compared to the lowest category LTE user equipment was made through release 12. Release 13 is expected to contribute another 50% cost reduction [75]. This is achieved in part by decreasing device bandwidth to 1.4 MHz and output power to 20 dBm.

A different approach in release 13 is narrow band Internet of Things (NB-IOT). It has similar goals with respect to power usage, range, and device complexity. It uses a bandwidth of 200 kHz which offers increased deployment agility but at the expense of decreased data rate and increased delay. Note that NB-IOT may be appropriate for 5G migration, supporting large numbers of inexpensive devices [54].

Latency Reduction: Latency (delay) is a very important system performance metric. Latency in LTE is less than 10 ms. For 5G, latencies of 1 ms is a key requirement. Reduced latency is crucial for good TCP/IP throughput. Latency reduction techniques are expected to be part of release 14. See [54, 75] for a discussion of such techniques.

Vehicle Communications: A number of novel and significant applications are possible with vehicle oriented communications. These include vehicle safety, traffic control, telematics, and infotainment. A number of studies were conducted by 3GPP on the use of LTE networks to provide connectivity between vehicles (V2V), between vehicles and road infrastructure (V2I), and between vehicles and pedestrians (V2P) or other mobile users. Collectively this functionality is called LTE V2X (V2X being vehicle to anything) [75].

There are other wireless protocol projects for vehicles such as IEEE WAVE and IEEE 802.11p but LTE V2X is attractive because of its ability for wide area coverage, connectivity to mobile users, and (spectrally) efficient V2I broadcast services. Use cases of V2X services and their requirements are discussed in release 13 [54]. The performance of some LTE solutions for supplying V2V, V2I, and V2P services will be the subject of a study with the aim of describing enhancements to appear beginning in release 14 [75].

Superposition Coding: This is the use of non-orthogonal transmissions with no spatial separation on the downlink [75]. As an example consider a downlink transmission to a user at the cell boundary and one near the cell center. These are transmitted with the same beam. The cell boundary user would usually be given a larger transmit power and so its interference to the cell center user can be cancelled before decoding the relevant signal at the cell center user. Release 13 described performance improvements and specifications which may appear in release 14.

Multimedia and Multicast: Mobile streaming video will be a major portion of network traffic in the years to come. Deployment of eMBMS (evolved multimedia broadcast multicast service) or "LTE broadcast" is of much interest as a means to deliver download and streaming content to multiple users. Improvements to eMBMS will be of use in applications such as linear TV, video

on demand, live, over the top content, and smart TV [54]. Some enhancements may be included in release 14 and there may eventually be a dedicated carrier for eMBMS.

4.6 Conclusion

This chapter has presented a system level description of wireless technologies IEEE 802.11 WiFi, IEEE 802.15 Bluetooth, and Long Term Evolution (LTE). The driving force in the development of new and novel protocols and technology in this area is the continued growing demand for increased data rates and communication flexibility. This growth in demand is likely to continue for some time.

Chapter 5
Multiprotocol Label Switching (MPLS)

5.1 Introduction

The idea behind Multiprotocol Label Switching (MPLS) is simple. In datagram switching each packet is treated as an independent element by routers. If one implements a Differentiated Services architecture, each independent packet is treated according to a policy specifically for its service class. In MPLS, on the other hand, the basic unit is a "flow." Packets belonging to a given source destination flow within an MPLS cloud of a network have a label(s) appended to the packet that is used by routers along a specified path to relay the packet from router to router.

In a sense the MPLS flow is similar to the virtual circuit concept used for such technologies as ATM. The technology of MPLS allows a certain degree of quality of service (QoS) for each flow and simplifies packet forwarding by routers (speeding their operation through the use of simple table lookups based on labels). Virtual private networks with good security can be set up using MPLS. Enhanced traffic engineering and more than one protocol (hence the "M" in MPLS) are also supported by MPLS. For instance, MPLS can carry both IP packets and ATM cells.

Because of its flexibility and elegance, MPLS has been very successful. It has been used in areas that were not originally anticipated and has been extended for use with different technologies. It has been used to implement virtual private networks (VPNs), to implement Border Gateway Protocol (BGP) free core networks, to provide traffic engineering for Internet Service Providers, for telecom providers through its MPLS-TP version, and in Pseudowire to transport layer 2 frames such as Ethernet [163].

Historically, several proprietary antecedents of MPLS were pushed by router vendors (CISCO, IBM, and others) in the 1990s. The IETF (Internet Engineering Task Force, see www.ietf.org) started working on a standardization effort in 1997 with the initial standard release in 2001 [11, 144].

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_5

5.2 Technical Details

The technology of MPLS has its own language. A label switched router is a router that follows the MPLS protocols. A set of routers or nodes that are adjacent to each other and form a single administrative domain for routing purposes is an MPLS domain. A router providing connectivity between a domain and nodes outside the domain is an (MPLS) edge node. A router receiving traffic entering a domain is an ingress node. Naturally, a router handling traffic leaving a domain is an egress node.

A label switched hop is a single hop connecting two MPLS routers that uses MPLS labels for forwarding. A label switched path is a sequential path through a series of label switched routers that packets belonging to a flow transit. That flow is a forward equivalence class. See [144] for a more complete listing of MPLS terminology.

In Fig. 5.1 an MPLS label is shown both in exploded view and as part of an datagram packet. In the datagram packet, PPP is a field for a protocol for data link framing such as with the Point to Point protocol [149] and CRC is a cyclic redundancy code field. The label consists of a 20 bit label number, 3 bit quality of service (class) designator, an S bit for label stacking (to be discussed) and a time to live field of 8 bits.

Labels are used only locally between a pair of adjacent routers. Thus a router may see the same label number on a number of incoming links. The label in an arriving packet is used to do a table lookup to determine the next outgoing line for the packet as well as the next label to use. A forwarding protocol such as MPLS can be considered as being between layers 2 (data link) and layer 3 (network).

There are two ways of setting up a label switched path. Under hop by hop routing each label switched router chooses the next hop for a flow. Under explicit routing a single router such as the ingress or egress router chooses some (loose routing) or all (strict routing) of the label switched routers a flow traverses [74, 144].

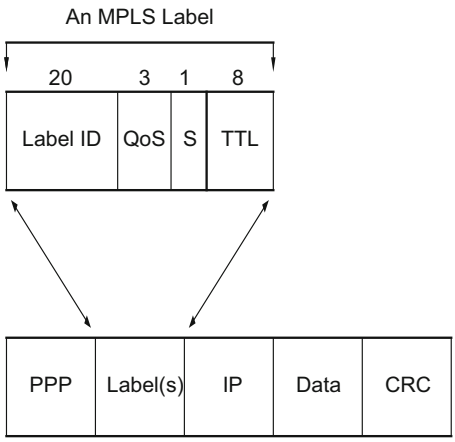


Fig. 5.1 An exploded view of an MPLS label embedded in an IP packet

Label switched paths in MPLS can be used as "tunnels." What is tunneling? Tunneling is a generic concept in networking involving one protocol's packets being encapsulated as data in another protocol's data field for transmission under the second protocol. For instance, suppose one has Ethernets in London and Paris. An Ethernet packet in London (consisting of header and payload) can be placed in the data field of a SONET frame for transmission to Paris. In Paris the Ethernet packet is removed from the SONET frame data field and placed on the Parisian Ethernet.

In the MPLS form of tunneling, a packet's path is specified by a label assigned by a label switched router. Naturally, the label switched path is set up first. Notice there is no need to specify the intermediate routers along the path. That is, the ingress node assigned label allows the packets to follow the (tunneled) path to the egress node.

Multiple labels may be used in the same packet in what is known as label stacking to allow the aggregation of flows. Multiple label switched paths can be aggregated into a single label switched path for some distance in an MPLS network. The labels are used and stacked in LIFO (last in, first out) form in a packet with the top label appearing first. For instance, imagine the use of three levels of labels. The top label could indicate the meta-flow to Europe belonging to some corporation. The middle labels could indicate the corporate traffic destined to particular European countries and the bottom label could indicate the corporate traffic for offices in different cities within each country. The S bit is set to "1" for the bottom label to indicate that it is the last label. One benefit of label stacking is smaller table size in routers.

Let's now take a look at the MPLS label fields. Although the 20 bit "label" field could theoretically support 2^{20} or a bit more than a million labels, actual implementations of label switched routers usually support a much smaller number of labels. Labels can be set up by the Label Distribution Protocol (LDP) or by LDP in conjunction with other protocols [74].

The QoS field with its 3 bits could in theory allow 8 classes of traffic. However, if one of the bits is used to indicate whether a packet can be discarded in the presence of congestion, only four classes of traffic can be supported with the remaining 2 bits. This is a limitation of MPLS [93].

Finally, the 8 bit time to live (TTL) field plays a similar role to the TTL field in IP packets. Decremented each time a packet makes a hop, the packet is deleted from the network when the TTL field reaches zero to prevent indefinite packet looping in the network.

5.3 Traffic Engineering

Normal routing in datagram based networks can lead to congestion and poor utilization of network resources. In such networks commonly used routing algorithms such as OSPF (Open Shortest Path First) will route packets along the shortest path. This can lead to congestion on shortest paths and under-utilization on longer paths. If there are multiple shortest paths one can use the ECMP (Equal Cost Multipath)

option of OSPF [95] but if there is only one shortest path ECMP is not effective. While one might consider manipulation of link costs/metrics this is not practical for large networks [166].

The process of evenly balancing traffic across a network to avoid congestion and under-utilization and to allow the maximum amount of traffic to flow is called traffic engineering. In fact MPLS is well suited to allow traffic engineering because, in terms of congestion mitigation, it is conceptually easier to assign and/or reroute flows than using the indirect method of changing link metrics in a datagram network [78].

Constraint based routing can be used to computer generate optimal or near optimal routes using several performance metrics simultaneously. For instance, for purposes of achieving good quality of service one may want to maximize data rate and minimize average link delay and packet loss probability. Algorithms such as shortest path routing use only one metric and are not considered constraint based routing algorithms. For constraint based routing individual metrics may be combined to produce a single overall metric, in additive fashion, multiplicative fashion, or using a minimum function.

In fact, as Xiao and Ni point out [166], constraint based routing is a routing methodology and MPLS is a forwarding scheme. It does not matter to MPLS forwarding how routes are chosen. Routing and forwarding are in theory independent. However, this is a case where the sum is greater than the parts. The technology of MPLS allows constraint based routing to use label switched path traffic information on flows through an MPLS domain. The flow paradigm of MPLS is well suited to joint use with constraint based routing.

5.4 Fault Management

When physical links or routers fail it is desirable to reroute traffic around the fault in a short amount of time so connections are not lost. Temporally, this can be done in MPLS statically (with pre-established backup paths) or dynamically (as faults occur).

In terms of network topology, there are a variety of ways to reroute traffic with MPLS [16, 86]. Under a "global repair model" an ingress node is made aware of a failure on a path either through a reverse message from the failure site or through a path connectivity test. It then reroutes traffic that was being transported along the original path from the ingress node itself. Under a local repair model, rerouting is done at the point of failure around the fault. Finally under reverse backup repair, traffic flow is reversed at the failure point back to the ingress node where it is continued on alternate path to the destination.

As discussed in Marzo, there are trade-offs between the various fault restoration methods. Global repair can be relatively slow, particularly if a path continuity test is used to detect a fault. Reverse backup repair can also be relatively slow. Local repair can be faster. However, there are also differences between the schemes in terms of the amount of network resources that need to be utilized/reserved.

5.5 GMPLS

At some point it was realized that the MPLS switching paradigm could be carried over to a variety of switching technologies. That extension is called GMPLS (Generalized Multiprotocol Label Switching). The "generalized" in GMPLS comes from the fact that the MPLS protocol is indeed extended to other switching technologies. For instance, for optical (DWDM) networks the concept of a label can be represented by a color (or wavelength) of a single stream. The development of GMPLS has been aided by such groups as the IETF (Internet Engineering Task Force).

A key advantage of GMPLS is its ability to provide guaranteed QoS and traffic engineering [32]. What GMPLS specifically does is provide a common control plane for packet (cell) switching, circuit switching, SONET, DWDM (dense wavelength division multiplexing), and fiber switching devices [108, 151]. The architecture of GMPLS provides for protocol capabilities such as signaling (as in RSVP-TE, Resource Reservation Protocol - Traffic Engineering), routing (as in OSPF-TE, Open Shortest Path First - Traffic Engineering), link management (as in LMP, Link Management Protocol), and fault recovery [108].

The traditional hierarchy of IP/ATM/SONET/DWDM is evolving towards IP/GMPLS over DWDM and IP/MPLS over DWDM hierarchies. This will simplify engineering, reduce costs, and improve performance [77].

5.6 MPLS-TP

An off-shoot of MPLS is MPLS-TP (TP is Transport Profile). It is used by telecommunication carriers as a packet switched data network. It was developed as a collaboration of the IETF (Internet Engineering Task Force) and the ITU-T (International Telecommunications Union—Telecommunication Standardization Sector). The IETF and ITU-T groups come from different perspectives (i.e., internet and telecommunication technology, respectively) but were able to work together on MPLS-TP. Joint work started by the two groups in 2008 (http://www.wikipedia.org).

In general, MPLS-TP allows MPLS technology to create a packet transport network (PTN) that is defined by ITU-T [163]. It is an evolution of classic transport network technology (which uses SONET/SDH - see the next chapter) to be more suited to packet switching. Note SONET/SDH is TDM based, overhead for use with packets [84]. However, MPLS-TP has features of that classic technology such as high availability, QoS, and the "operational look and feel" [163].

To reach this goal, MPLS-TP uses only some of the (relevant) features of MPLS and adds new ones largely involving OAM (operations, administration, and management), network management, and survivability techniques. See [163] for a detailed discussion of features.

Chapter 6
Optical Networks for Telecommunications

Of all physical transmission media, fiber optics has the highest data carrying capacity. Networks using fiber optic links, including telecommunication networks and the Internet, are well established and continue to evolve. In this chapter fundamental optical network technologies are discussed.

6.1 SONET

SONET (Synchronous Optical Networking) is a popular standard for fiber optic voice and data transmission. It was developed originally by Bellcore, the research and development arm of the local American phone companies in the late 1980s [140]. It was meant to be a standard for fiber optic connections between telephone switches. However, it was a technology at the right place, at the right time and has been extensively used over the years for telephone trunk transmission and internal corporate and governmental traffic. More specifically, it was developed at about the time that there was an interest in providing broadband integrated services digital network (B-ISDN) technology. After its creation it was used to carry ATM [1] traffic, IP packets, and Ethernet frames (http://www.wikipedia.org).

SONET, when it was developed, took into account B-ISDN technology, political and international compatibility concerns. The SONET architecture is elegant and took advantage of LSI and software advances at the time. Development has continued over the years with the introduction of higher and higher standardized data rates.

[1] Asynchronous Transfer Mode (ATM) is a packet switched technology developed originally by the telephone industry and used for a number of years in the Internet. It uses virtual circuits and a 424 bit fixed size packet [79, 104, 121, 123, 148].

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_6

Table 6.1 Some SONET rates

Acronym	Gross rate
STS-1/OC-1	51.84 Mbps
STS-3/OC-3	155.52 Mbps
STS-12/OC-12	622.08 Mbps
STS-48/OC-48	2.48832 Gbps
STS-192/OC-192	9.95328 Gbps
STS-768/OC-768	39.81312 Gbps

Table 6.2 Virtual tributary rates

Acronym	Data rate
VT1.5	1.728 Mbps
VT2	2.304 Mbps
VT3	3.456 Mbps
VT6	6.912 Mbps

A typical SONET data rate is abbreviated as STS-n/OC-n where $n = 1, 2, 3 \ldots$ The "STS" indicates the electrical interface and the "OC" indicates the optical interface. The STS-1/OC-1 rate is 51.84 Mbps. Any other STS-n/OC-n rate is n times faster than 51.84 Mbps. For instance, STS-3/OC-3 is at 155.52 Mbps. In fact, STS-3/OC-3 is the lowest SONET rate used in practice. Table 6.1 indicates some of the various SONET rates.

Lower rates, known as virtual tributaries, are also available. For instance, virtual tributary 1.5 (VT1.5) is at 1.728 Mbps. Some virtual tributary rates are indicated below (Table 6.2).

Note that VT1.5 is compatible with the T1 rate of 1.544 Mbps and VT2 is compatible with the European version of T1 rate of approximately 2.0 Mbps. SONET is used in the USA and Canada. Its cousin, SDH (Synchronous Digital Hierarchy), is used elsewhere. In fact SONET is considered a version of SDH although it was created first (http://www.wikipedia.org).

6.1.1 SONET Architecture

SONET data is organized into tables. For STS-1/OC-1, the byte table consists of 9 rows of bytes and 90 columns of bytes (Fig. 6.1). As shown in the figure, the first 3 columns hold frame overhead and the remaining 87 columns hold the payload. Some additional overhead may appear in the payload. Each byte entry in the table holds 8 bits. If digital voice is being carried, the 8 bits represent one voice sample. Uncompressed digital voice consists of 8 thousand samples/s of 8 bits each (or 64 Kbps). Thus, the SONET STS-1/OC-1 frames are generated at a rate of 8K frames/s.

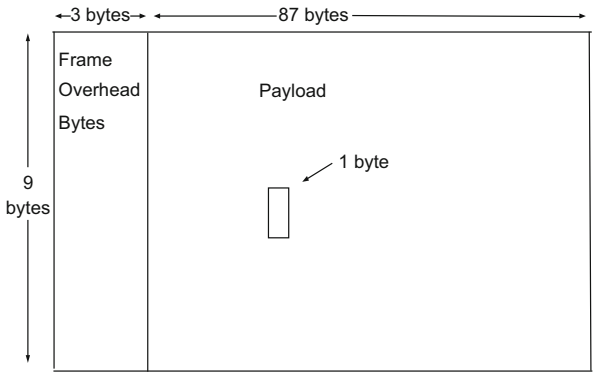
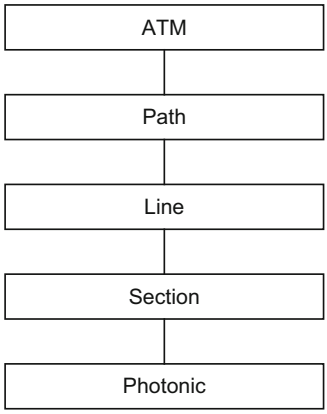


Fig. 6.1 An STS-1/OC-1 SONET frame

Fig. 6.2 SONET protocol stack



The protocol layers for SONET go by the names of path, line, section, and photonic (see Fig. 6.2 where ATM is being carried over SONET). The functions of the layers are [104]:

- Path = End to end transport as well as functions including multiplexing and scrambling.
- Line = Functions include framing, multiplexing, and synchronization.
- Section = Functions include bit timing and coding.
- Photonic = Physical medium.

Overhead of each type appears in an STS-1/OC-1 frame as illustrated in Fig. 6.3.

Note that the start of a payload is indicated by a pointer in the line overhead.

The SONET system layers can be viewed in terms of a box type diagram as in Fig. 6.4.

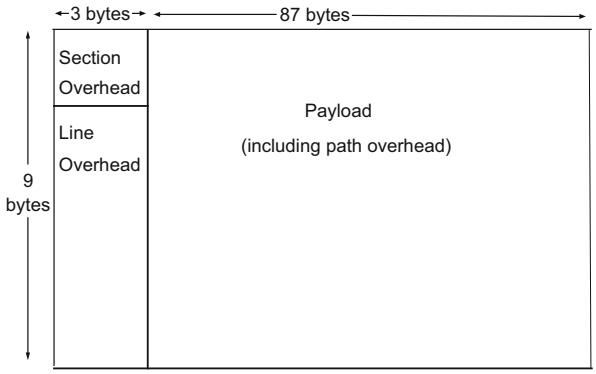


Fig. 6.3 An STS-1/OC-1 SONET frame with overhead indicated

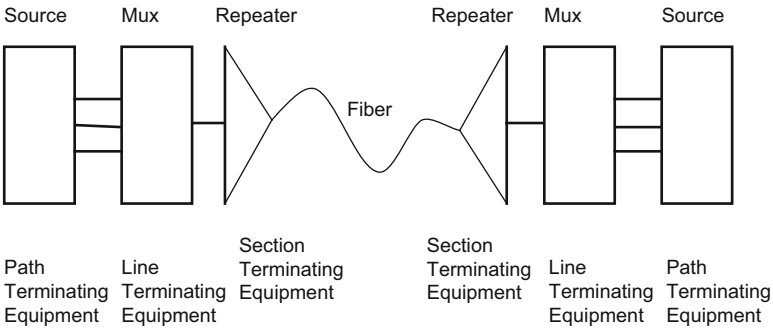


Fig. 6.4 SONET system diagram

There are two major configurations for the SONET opto-electronic interface at a node. If the fiber starts/ends at the node, one says one has a SONET Add Drop Multiplexer (ADM) in terminal mode. The ADM allows signals to be tapped off or on the fiber. Alternately, one may have fiber passing through the node. That is, a fiber enters from the east, for instance, is converted to an electrical signal, signals are tapped off and inserted and a new fiber leaves to the west. This is called a SONET ADM in add/drop mode.

6.1.2 Self-Healing Rings

One of the reasons for the widespread use of SONET is it allows a variety of topologies. The most useful of these are ring topologies. While one can implement linear add/drop networks (see Fig. 6.5), if a fiber in these is cut or an opto-electronic transceiver fails, one loses connectivity.

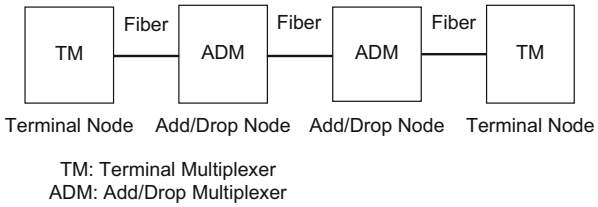
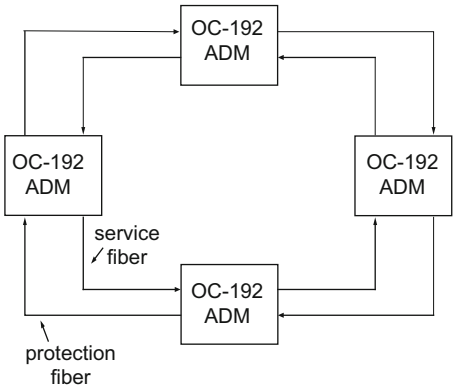


Fig. 6.5 A SONET linear network

Fig. 6.6 A SONET ring network



Typically, rings are implemented with multiple service and protection fibers (see Fig. 6.6). If a service fiber path fails, a protection (backup) fiber is switched on its place. The number of protection fibers can be less than the number of service fibers if one is willing to live with less redundancy. Also, if all of the fibers between two adjacent nodes are lost, with a sufficient number of protection fiber rerouting can keep a logical ring in place until repairs are made.

A competitor, at least for data traffic, to SONET is 10 Gbps Ethernet [90].

SONET rates have increased over the years but not by enough for fiber to reach its full potential unless a second technology, wavelength division multiplexing, is also used. This WDM technology is discussed in the next section.

6.2 Wavelength Division Multiplexing

To boost capacity in a fiber network one can increase the number of cables, boost the bit rate of each channel or multiplex signals on a single fiber. Increasing the number of cables may not be feasible for economic and practical reasons. Increasing SONET rates, for instance, may not be viable. So the economic solution in many cases is to multiplex signals on a single fiber [98].

In wavelength division multiplexing (WDM), special multiplexing at either end of a fiber can put multiple optical signals in parallel on a single fiber [125]. Thus,

instead of carrying one SONET OC-192 signal at about 10 Gbps, a single fiber might carry 40 OC-192 signals or 400 Gbps of traffic. Even terabit range (1000 Gbps) capacity is possible. Current systems can have up to about 80 channels and possibly more. With each signal at a different optical frequency, WDM is essentially a frequency division multiplexing scheme.

6.2.1 History and Technology

Wavelength division multiplexing was first used in the early 1980s with two optical streams, one at 850 nm and one at 1300 nm. A simple fused coupler allowed the injection of both signals into a multi-mode fiber (see Chap. 1). At the receiver a silicon detector sensitive to 850 nm light and a germanium detector sensitive to 1300 nm light were used. Filters suppressed unwanted wavelengths.

In the late 1980s single mode fiber was in use. Fused couplers that could separate 1300 and 1550 nm streams with cheap hardware were available. However, fiber optimized for 1300 nm light and fiber optimized for 1550 nm light are different. Generally fiber optimized for 1300 nm was used for local loops. Fiber optimized for 1550 nm was used for long haul and submarine cables [98].

Also in the late 1980s fiber optic amplifiers became available and WDM was practical in the fiber optic amplifier's operating region of 1520–1560 nm. Soon four signal systems spaced by 10 nm were demonstrated. The input of a WDM system is a simple coupler that multiplexes inputs into one fiber. Couplers are available to multiplex 2, 4, 8, 16, 32, and 64 inputs.

The de-multiplexer is a bit more complex. Light from the fiber is collimated into a narrow beam of light. The lights impinges on a grating which, working like a prism, sends light at different WDM frequencies off at different angles where optics collects the individual beams and focuses them into separate output fibers [98].

Most DWDM systems are designed to operate in the fiber optic amplifier window of 1520–1560 nm. Fiber can now be made with less sensitivity to absorption in the OH bands (1400 and 1600 nm). This increases the range of DWDM systems as long as wider range fiber optic amplifiers are developed.

The telecommunications history of large scale WDM goes back to a fiber glut in the USA that existed prior to 1995. After the Bell System divestiture in the 1980s, the competitors in long distance phone service had financial limitations so relatively low numbers of fiber per path were laid (usually about 16 fibers). But by the end of 1995, the interexchange fibers of the long distance carriers were nearing exhaust. In 1996, 60% of the AT&T network, 84% of the MCI network, and 83% of the Sprint network were fully lit [125].

About this time WDM technology became practical. This technology included distributed feedback lasers needed to produce the monochromatic output of WDM, filters to separate signals which are closely packed in frequency and optical amplifiers. In particular, Erbium Doped Fiber Amplifiers (EDFA) allowed the amplification of optical signals without intermediate electronic conversion.

In 1994, Pirelli/Nortel introduced 4 channel systems and IBM introduced a 20 channel system. Cienna followed with a 16 channel system in 1996. By 1997/1998, 32 and 40 channel systems were being produced. It should be noted that Cienna's successful WDM products led to a very successful public offering and 196 million dollars in first year revenue (the fastest in corporate history at that point).

Systems using WDM can carry a variety of traffic such as SONET, Ethernet, and cable TV signals.

Depending on the wavelength assignment WDM systems may be either conventional/coarse (CWDM) or dense (DWDM). More channels are accommodated with denser channel spacing with DWDM compared to CWDM (http://www.wikipedia.org). Coarse WDM systems operate over the range of 1260–1670 nm with signals spaced by 20 nm each. But it is only feasible if high water absorption bands are not present in the fiber [98].

At first, WDM technology was used in long distance networks but as its costs decreased, metropolitan area network usage followed.

Tunable lasers have been introduced as a way of providing backup. That is, rather than having 32 fixed wavelength spares for a 32λ system, one tunable laser provides protection against the most likely case, a single bad fixed wavelength laser.

6.2.2 Switching

Three generic switching technologies can be used to carry IP traffic over WDM [28]. These are optical circuit switching, packet switching, and optical burst switching.

For optical circuit switching to be efficient, like circuit switching in general, the length of transmissions needs to be significantly greater than the circuit setup time. This is often not true of bursty data traffic. It has also been shown that the circuit establishment problem is in general NP hard. That is finding an optimal solution is computationally intractable though sub-optimal heuristic solution techniques can be used.

The technology for buffering and packet processing, used by packet switching techniques, is not yet mature in the optical area making practical, cost efficient, systems not possible at this time.

The proponents of optical burst switching feels it has the advantages of both circuit and packet switching. The burst is the fundamental element of optical burst switching. It can be thought of as a variable length packet with a control header and the payload to be carried. An optical burst switching system operates over edge and core routers. Edge (of the network) routers can be either ingress or egress routers. A variable length burst is assembled at an ingress router from multiple IP packets, possibly from multiple hosts. The burst is transported over the core network and its routers. At an egress router the burst is disassembled.

6.3 Optical Transport Networks

Optical Transmission Networks (OTN) are described in a suite of standards including ITU G.872 and G.709. They offer high speed data transport as well as the ability to carry a mix of lower speed network traffic. Optical transmission networks were created in the 1990s. While it is not a new technology, it is relatively new to many in the optical industry. Unlike SONET/SDH, OTN is designed to take advantage of wavelength division multiplexing for optical networks (FS.COM 2016, [141], http://www.wikipedia.org).

The initial motivation for OTN was to create an integrated infrastructure for multiple services. This includes SONET/SDH, Ethernet, the Internet Protocol (IP) etc. Among the advantages of using OTN are [141]:

- Improved wavelength operations.
- Multiple service classes supported by a common container. OTN can be called a "digital wrapper."
- A common multiplexing hierarchy (i.e., OTU1 at 2.5 Gbps, OTU2 at 10 Gbps, OTU3 at 40 Gbps, and OTU4 at 100 Gbps). In addition there is scalability of switching.
- A coding gain (reduced error probability) through the use of Forward Error Correction (FEC). The use of forward error correction improves the optical signal to noise ratio by 4–6 dB. This allows longer spans and less regeneration (FS.com 2016).
- Additional levels of Tandem Connection Monitoring (TCM).

One of the good points of OTN is it includes operation, administration, and maintenance (OAM) functions for multiwavelength optical carrier systems. It has the network management ability of SONET/SDH but for wavelengths. Its frame size is flexible and multiple frames of data can be "wrapped" into a single unit. This single entity can be efficiently managed with its smaller overhead (FS.com 2016).

The OTN (G.709) standard involves a frame that wraps data packets much as a SONET frame does. There are six "layers" to this format (MetaSwitch undated, FS.com 2016) [141]:

1. OPU (Optical Channel Payload Unit): Client data is mapped into the OPU payload as well as a header (overhead) describing the type of data involved and mapping format. The OPU is similar in a general sense to the path layer of SONET/SDH.
2. ODU (Optical Data Unit): ODU overhead includes optical path level monitoring, claim indication, protection switching bytes, and embedded communication channels. The ODU is the payload that is groomed and switched in the OTN network. The ODU plays a similar role as the line overhead does in SONET/SDH.
3. OTU (Optical Transport Unit): The OTU is an optical interface/port for different OTU rate streams. The OTU also includes performance monitoring for the optical layer, alarm indication, and a communication channel. It also adds the Forward Error Correction (FEC) coding. The OTU is similar to the section overhead in SONET/SDH.

4. OCh (Optical Channel): The end to end optical path at some wavelength.
5. OMS (Optical Multiplex Section): This handles fixed wavelength DWDM connecting OADMs (optical add/drop multiplexers).
6. OTS (Optical Transport Section): Each OMS is given an overhead channel on the same wavelength to create the OTS. Together the OTS, OMS and OCh channel overheads can be used to evaluate transmission channel quality and for fault detection.

6.4 Flexible/Elastic Optical Networks

A flexible/elastic optical network is able to allocate spectrum bandwidth to light paths in response to the bandwidth needs of users/clients. The spectrum is broken down into narrow optical frequency slots and connections are given different numbers of optical frequency slots. This paradigm allows the creation of "super-channels," comprised of densely packed sub-channels, and makes available tunable bit rates of tens of gigabits per second up to the terabit per second range. Several terms are used for such "flexible" networks in the literature including "elastic," "flexgrid," "flexigrid," "tunable," and "gridless" [29, 152].

The technology of flexible optical networks allows a significant improvement in network utilization compared to what is possible in DWDM optical networks. A number of parameters are tunable in flexible optical networks such as optical data rate, modulation format, and wavelength spacing between channels. These parameters are fixed in current optical networks.

6.4.1 Numerical Examples

In terms of some numerical examples [29, 152], under a classic WDM based optical network channels are spaced 50 or 100 GHz apart as specified by ITU-T (International Telecommunications Union - Standardization Sector) standards. Such a channel will be under-utilized if it carries too little traffic to fill a 50 GHz slot.

For flexible optical networks suppose slots are 6.25 GHz wide (i.e., center frequencies are 6.25 GHz apart). Then one optical signal may fill four such slots for a total bandwidth of 25 GHz and another signal may fill 6 slots for a total bandwidth of 37.5 GHz. Note that even numbers of slots must be allocated to a signal as a pair of 6.25 GHz slots must be allocated about the central frequency each time. Note also that using orthogonal modulation one can allow some overlap of signal spectrum increasing utilization efficiency.

Finally, note that lower rate signals can be multiplexed into higher rate signals. A 40 Gbps optical bandwidth may be partitioned into 5, 15, and 20 Gbps sub-signals. Or three 40 Gbps optical bandwidths may be combined with optical OFDM (orthogonal frequency division multiplexing) into a 120 Gbps "super-channel" [29].

6.4.2 Network Characteristics

Some of the characteristics of flexible optical networks are [29]:

- Bandwidth Segmentation: There is inherent bandwidth wastage in DWDM networks when only partial bandwidth is required (in DWDM one must allocate the full channel bandwidth). Flexible optical networks allocate "just enough" bandwidth to meet user demand. This more efficiently uses network resources.
- Bandwidth Aggregation: Lower bandwidth requirements can be contiguously combined into super-wavelength optical paths.
- Energy Savings: When traffic is light, energy can be conserved by turning off some of the OFDM sub-carriers.
- Network Virtualization: Virtual links can be supported by OFDM sub-carriers.

6.4.3 Routing and Spectrum Allocation

The Routing and Spectrum Allocation (RSA) problem in flexible optical networks involves finding appropriate routes for source/destination pairs and allocating appropriate spectrum to each requested light path.

The problem is equivalent to the Routing and Wavelength Assignment (RWA) problem for WDM optical networks. The RWA problem involves defining light paths for each connection request by choosing a suitable route and assigning a wavelength. In RWA the same wavelength is used for a connection's end to end path (this is the wavelength continuity constraint). In the RSA problem a set of contiguous spectrum slots is assigned to a connection in place of the wavelength set of RWA (spectrum continuity constraint).

Both RWA and RSA are NP hard computational problems. However, progress can be made by dividing the RSA problem into separate routing and spectrum allocation sub-problems [29].

6.5 Passive Optical Networks

Access networks connect provider (central office) facilities to multiple users. Access networks may be wired or wireless [135]. In this section Passive Optical Networks (PONs), a popular optical access technology, are discussed. They are popular due their attractive cost, service transparency, energy impact, and good security. Passive optical networks use low power passive elements which do not need power in the fiber distribution networks [2]. Ranges for access networks can be hundreds of meters up to 20 km or so (Fig. 6.7).

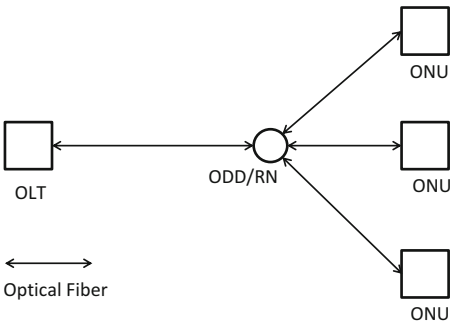


Fig. 6.7 Generic architecture of a passive optical network. Downstream is *left to right* transmission and upstream is *right to left* transmission

A generic architecture for a passive optical network is shown in the accompanying figure [135]. The Optical Line Terminal (OLT) is on the provider's premises (central office) supplying and receiving signals to/from the end users. The downstream data wavelength is λ_d and the upstream data wavelength is λ_u. The Optical Distribution Node (ODD) or remote node (RN) demultiplexes the downstream signal to the end users and multiplexes the upstream signals from the end users to the OLT. The end users' Optical Network Units (ONU) are located near or on customer premises.

There are a number of transmission technologies that can be used in connection with PONs. Three popular ones discussed below [2, 135] are Time Division Multiplexing (TDM), Wavelength Division Multiplexing (WDM), and Orthogonal Frequency Division Multiplexing (OFDM).

6.5.1 Time Division Multiplexing PON

A TDM passive optical network uses a passive power splitter/combiner at the ODD/RN. The number of ONUs connected to a single RN is up to 32, 64 or 128 ONUs. The central office uses N time slots for the N users ($ONU_1, ONU_2, \ldots ONU_N$). For bi-directional TDM, which is possible, optical circulators are used at the central office OLT and end user's ONU to separate distinct downstream and upstream signals. A problem with TDM is the N end users share the bandwidth so each end user receives only $1/N$ of the bandwidth.

Time division multiplexed PONs come in three standard varieties [135]:

- Ethernet PON (EPON): 1.0 Gbps symmetrical operation.
- Broadband PON (BPON): 622/155 Mbps downstream/upstream rates.
- Gigabit PON (GPON): 2.5/1.25 Gbps downstream/upstream rates.

A TDM PON is not highly secure because of the shared information infrastructure and possible eavesdropping. A number of approaches to increase TDM capacity have been researched [2].

6.5.2 *Wavelength Division Multiplexing PON*

Under this approach a separate wavelength is used to connect the OLT to each ONU in each direction. Thus there are wavelengths $\lambda_{d1}, \lambda_{d2}, \ldots \lambda_{dN}$ in the downstream direction and wavelengths $\lambda_{u1}, \lambda_{u2}, \ldots \lambda_{uN}$ in the upstream direction.

The full data rate of a wavelength is available to each ONU, an improvement over TDM PONs. The ODD/RN splits the wavelengths coming from the central office to each ONU end user and combines the upstream optical signals from the end user ONUs into a multiplexed signal for transmission to the central office. Grating type components are used in the ODD/RN. On the downside, WDM PONs require specialized components for each end user which raises cost. Techniques to mitigate the cost element are discussed in [135].

6.5.3 *OFDM PON*

Following the generic PON architecture, Orthogonal Frequency Division Multiplexing (OFDM) uses downstream wavelength λ_d and upstream wavelength λ_u. The central office OLT creates orthogonal sub-carriers using the Fast Fourier Transform algorithm. This is OFDM Modulation. A number of orthogonal sub-carriers are assigned to each ONU. The downstream multiplexed signal is split at the ODD/RN into N signals, one for each ONU. An ONU obtains its allocated sub-carriers and converts them to time signals using the inverse Fast Fourier Transform algorithm. This is OFDM modulation.

For the upstream traffic the end user ONUs create sub-carriers which are modulated to wavelength λ_u. These are combined at the ODD/RN and sent over fiber to the central office [135].

There are some cost advantages to the use of OFDM PONs. However, OFDM PONs need receivers based on high data rate DSP and FPGAs. There are also noise issues [2].

Note that TDM, WDM, and OFDM are not the only technologies that can be used with PONs and hybrid schemes are another possibility.

6.6 Orbital Angular Momentum

Light traveling through space in a spiraling fashion has a kind of momentum called orbital angular momentum (OAM) [161]. This is surprising but true. As an example a solar sail can accelerate through space moved by sunlight. Circularly polarized light has spin angular momentum and can induce a torque in a microscopic object it hits. In 1992 it was found by Allen and Woerdman that a spiraling beam has orbital angular momentum. It can cause a tiny object to rotate or move in an orbit about the beam's center.

In terms of communications it should be possible to separately modulate (i.e., impart information to) each of many spiraling beams and superimpose them on each other. Within practical limitations the beams shouldn't interfere with each other. This is true of both radio and optical beams. The promise, if this can be made to work at a reasonable cost, is a big boost in multiplexed capacity.

Work has been done for such systems for radio frequency links, free space optical links, and fiber optic communications. Work has progressed the furthest for the first two situations.

There are challenges with using OAM with fiber. When OAM light is sent through fiber, bends in the fiber and temperature changes may alter the light's phase profile so some energy presents itself as a wave with a different amount of OAM. Using special fibers (e.g., vortex fiber) to mitigate this is one possibility. If successful it is likely that OAM based fiber will be used over shorter distances such as in data centers and high performance computer centers. Also transmission/reception equipment that is cost effective and efficient needs to be developed [161]. It should be noted that a photon can have many potential values of OAM. This has implications for increasing the capacity of a quantum link.

Chapter 7
Software-Defined Networking

7.1 Introduction

Software-defined networking (SDN) is an architecture and design strategy for designing programmable networks. The programmability makes it far easier, than is possible with the current Internet architecture, to implement new features, do network monitoring, and provide scalability.

7.2 Classic Internet Architecture

A generic way of describing the functionality of computer networks is in terms of three "planes." Working our way upward from lower level functionality to higher layer functionality (Fig. 7.1):

Data Plane: Consists of network devices such as switches and routers that transmit/forward packets.
Control Plane: Involves protocols to create/update forwarding tables in the data plane devices.
Management Plane: Here are software services such as SNMP (Simple Network Management Protocol) tools. These are used to manage/configure control functions.

A major problem with the classic Internet architecture is that the control plane and data plane are vertically implemented in an integrated manner within each individual device (i.e., switch) as in Fig. 7.2a.

This is a hardware centric approach. With this approach implementing new high level policies requires each individual network device to be configured using primitive and sometimes vendor dependent commands. Such high level policies could

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_7

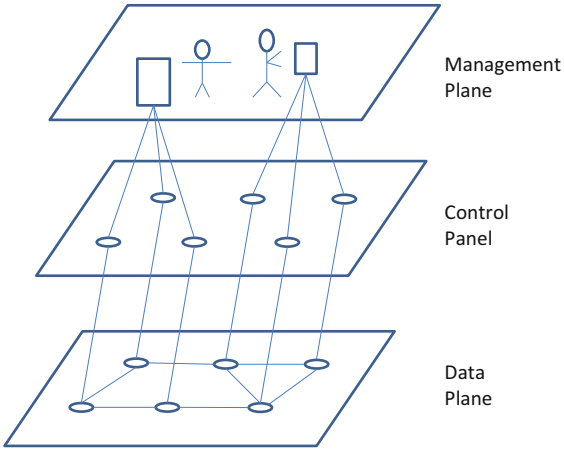


Fig. 7.1 Three network layers

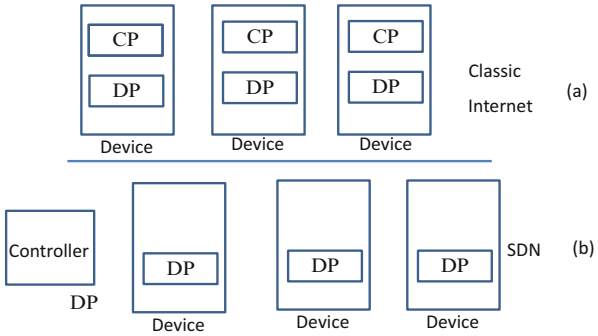


Fig. 7.2 (**a**) Classic and (**b**) SDN network architectures. Here CP is control plane and DP is data plane

be a new protocol, changes to an existing protocol or new services. The present Internet (IP network) has no automatic reconfiguration and response mechanisms. Correcting protocol faults is a related problem.

Examples of the problems that can occur with the current Internet architecture include the far from finished transition from Internet protocol version 4 (IPv4) to IPv6, and the 5–10 years it takes for a new routing protocol to be created, evaluated, and implemented in the field [66]. As Farhady puts it, "...innovation and research is costly under the current condition of hardware centric networking [43]". It would be far simpler and much less time consuming to support new protocols and services if the Internet could be programmed in a flexible and responsive manner.

7.3 SDN Architecture

Software-defined networking is a concept that is receiving increasing attention as a means to create a more flexible network architecture. Instead of vertically integrating the control and data planes in the same device, the control capabilities of the control plane are external to the data plane's switches and routers (see Figs. 7.2b, 7.3). The switches are now single "forwarding devices" [66]. An important point is that while control logic is centralized, it is logically centralized, not physically centralized. Physical centralization of functionality has well-known problems of scalability and single point of failure reliability.

From Fig. 7.3, the applications can utilize the decoupled control and data plane to achieve goals such as network monitoring. Applications communicate with the controller through the "northbound" interface (API), which has not been standardized to date. Controllers can communicate with each other through an interface which also has not been standardized yet. The controllers communicate with the SDN functional switches through the "southbound" interface. OpenFlow is the most popular such interface, though not the only one (see [66] for a list of OpenFlow hardware and software).

In short, software-defined networking has been defined by the Open Networking Foundation (in their SDN Definition) as:

> ... *"an emerging network architecture where the network control is decoupled and separated from the forwarding mechanism and is directly programmable."*

More specifically, an (OpenFlow) switch contains one or more forwarding tables under the control of a centralized controller. Programmability thus resides in the control plane. These forwarding tables contain rules that control packets (i.e., for forwarding, dropping, modifying). An OpenFlow switch can act as a switch, router, firewall, load balancer, traffic shaper, or network address translator (NAT). This eliminates the need for middle boxes which implement some of this functionality and are often placed in classical IP networks.

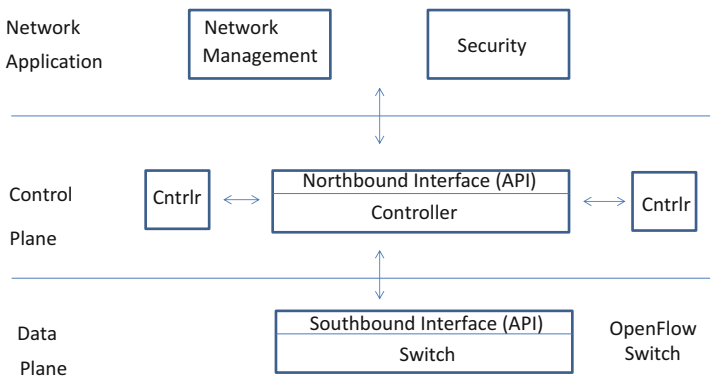


Fig. 7.3 Software-defined network architecture

Kreutz et al. [66] describe four elements of the SDN architecture:

1. The control and data planes are separated—switches are simple forwarding devices.
2. Flow based forwarding is used rather than the destination based forwarding. This allows a lower packet processing overhead as opposed to routing individual packets. A "flow" is a packet sequence traveling from a source to a destination (used, for example, in MPLS). Each packet in a flow obeys identical service rules at a forwarding device. The flow concept allows a unified implementation of different types of network functionality (routers, switches, middle box functions) allowing historically new flexibility. One is only limited by the capabilities of flow tables.
3. Control logic no longer resides in switches and similar devices but instead resides externally in SDN controllers or, if you like, in the Network Operating System (NOS). The NOS can be thought of as a software platform running on commodity servers that allows the generic software base and software resources to program forwarding devices (through flow tables). This programming takes a "logically centralized network view" [66]. In this way it is like a traditional computer operating system.
4. Software applications on the top of the network operating system can program the network. The NOS interfaces with the low level data plane devices.

The SDN approach has benefits in terms of the logical centralization and control and also in terms of the separation of the control and data planes. As was mentioned, control is logically centralized in SDN [66]. This provides some real advantages. Changing network policies is less likely to lead to errors and is more tractable using higher level languages and software instead of doing device specific reconfigurations. A (logically) centralized controller makes the creation of sophisticated network functions and applications much more doable in a reasonable amount of time. It is also easier for a controller to react to unexpected network state changes and continue operation of policies.

In terms of the separation of the control and data planes, it is much simpler to program applications and services under SDN than with the classic Internet architecture because the programming abstractions of the control program and programming languages can be shared. Policy decisions are more consistent because of the global network view of the logically centralized controllers. Forwarding devices can be reconfigured from any point in the network so that location of functionality is no longer a concern. Function integration is also simpler in the high level programming environment of SDN (for instance, it is easy to make functions sequential).

7.4 Development of SDN

A number of ideas used in SDN have appeared previously in other networking work [43]. There is a large number of such efforts.

Active Networks: In one version of "active networks," a packet contains a program fragment instead of data. The program fragment is executed by a node that the packet is in and actions are taken, depending on the data plane design, such as forwarding or dropping the packet. However, note that this concept emphasizes data plane programmability rather than control plane programmability (as SDN does) [43].

A second version of active networking does not change packet format but rather has an ability to download programs to switches. These programs detail how to process packets [66].

Open Signaling: The Open Signaling community, which grew in the 1990s, had a goal of making it possible to use open programmable interfaces. This would allow open access to switches and routers. Taking a telecom industry approach, this would allow third party providers to enter the telecom software market. Like SDN, OPENSIG delineates boundaries between area of functionality. In the case of OPENSIG these were network transport, network control, and network management.

Separation of Control Plane and Data Plane: There have been projects that sought to separate the control and data plane because of advantages of simplicity and abstraction [43]. These include:

- Network Control Point: This was a telephone industry effort to separate control and data plane signaling. It is perhaps the earliest implementation of the concept.
- Tempest: This was an ATM effort to allow co-existing virtual networks on a shared physical infrastructure. It proposed a software controller to handle forwarding devices.
- ForCES: An Internet Engineering Task Force (IETF) effort and standard providing an interface between control and data plane.

Centralized Network Control: Efforts with this goal include Routing Control Platform (RCP), Path Computation Effort (PCE), Soft Router, and Intelligent Route Service Control Protocol (IRSCP) [43].

7.5 OpenFlow

The popular OpenFlow architecture preceded the creation of the term software-defined networking. They are not identical. OpenFlow specifies an API (application programming interface) used to interface between a controller (in the control plane)

and switches (in the data plane). But OpenFlow is only one of the several available means of managing forwarding in a network wide manner[1] [43].

OpenFlow allows both hardware and software implementation of SDN. It is being integrated into commercial switches and routers [55].

Ethane was the forerunner of OpenFlow. In 2006 the SANE/Ethane project proposed a new enterprise network architecture. A centralized controller would control network policy and network control. A controller would make decisions on packet forwarding and there would be Ethane switches holding a flow table and connections to the controller.

OpenFlow provides a standard for information exchange between the control and data planes [13]. Forwarding devices hold one or more than one flow table plus an abstraction layer. The abstraction layer communicates with a controller in a secure manner. Flow entries consist of [13]:

Match Fields: These are matching rules used to match packets that enter the forwarding devices. The matching is based on information in the packet header, metadata, and/or ingress port.
Counters: Counters are used to hold statistics for individual flows. Such statistics include the number of received packets, number of bytes, and flow duration.
Actions: An action is a set of instructions that detail how to handle matching packets.
Table Miss Entry: This entry indicates what action to take if there is no match for a packet (i.e., table miss). Actions for a table miss may include packet dropping, continuing the matching operation, or forwarding the packet to the controller for further action.

A remote controller can modify table entries using the OpenFlow protocol. Starting with version 1.1, OpenFlow allows multiple tables as well as pipeline processing. "Hybrid switches" can have both OpenFlow ports and non-OpenFlow ports. The alternative ForCES protocol does not obey the same SDN model used by OpenFlow but can implement similar functionality [13].

7.6 Two Issues

Two important issues for SDN are controller scalability and security.

Controller Scalability: Astuto [13] points out that the original controller of Ethane run on a desktop could process 11,000 flow requests per second (with 1.5 ms latency). A 2012 study of an emulated network of up to 256 switches and 100,000 endpoints with OpenFlow implementation of controllers processed at least 50,000 flow requests per second per switch. Thus scalability to large numbers of flows should be tractable.

[1] Alternatives include JunOS SDK and Cisco ONE [43].

Security: Besides the usual places for attacks on networks (routers, switches) the emergence of SDN provides targets such as controllers, virtual infrastructure, and in the OpenFlow protocol on devices using OpenFlow [55]. A discussion on security appears in Hu.

7.7 Standards

Lists have been published of OpenFlow controllers [43, 59] and also of debuggers and test beds [43].

Many standards groups are making OpenFlow standardization efforts [66]. These include efforts by the Open Networking Foundation (ONF), the Internet Engineering Task Force (IETF), the International Telecommunications Union Telecommunications Sector (ITU-T), the Broadband Forum (BBF), the Metro Ethernet Forum (MEF), the IEEE 802 LAN/MAN standards committee, the Optical Internetworking Forum (OIF), the Open Data Center Alliance (ODCA), the Alliance for Telecommunications Industry Solutions (ATIS), the European Telecommunications Standards Institute, and the 3GPP consortium.

Chapter 8
Networks on Chips

8.1 Introduction

In this chapter the smallest possible networks are examined: networks implemented on a single integrated circuit chip. There is a reason that such "Networks on Chips" (NOC) have become important and even essential to today's and the future's electronic technology.

That reason has to do with Moore's law [150]. Moore's law, which was valid for decades, held that through higher clock frequencies and continual miniaturization of chip circuits, computer processing power would double every 18–24 months. However, in recent years the operating frequencies of chips have not increased as Moore's law predicted. This is due to unsustainable power demands and constraints on thermal density. But transistor density has continued to increase.

Thus the solution adopted in recent years is to make use of parallel processing by including many (multi) "cores" on the same chip. This development came after systems on chips (SOCs) became popular, starting in the 1990s. Systems on chips integrate processors, video processors, graphics engines, embedded memory, I/O devices, and even analog components on a single chip [102, 150].

Classic on-chip communication architectures are buses and point-to-point (P2P) interconnection networks (Fig. 8.1). Both architectures have scalability issues though. Bus architectures can connect a few tens of cores with minimal cost. However, unless one is broadcasting, a bus is an inefficient medium as communication can only occur between two processors/components at one time. This inefficiency worsens as the number of processors is increased. There are also power consumption issues with buses. Finally, there is the significant capacitative load of bus drivers, which creates appreciable delay and energy use in interconnected wires [102].

Point-to-point interconnection, at least for a complete interconnection (every core directly connected to every other core), for N nodes has an $O(N^2)$ complexity. This

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_8

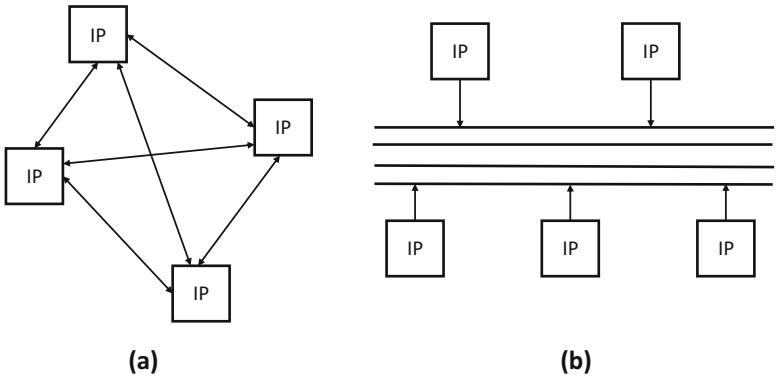


Fig. 8.1 Classic interconnection networks: (**a**) point-to-point interconnection and (**b**) bus interconnection

makes it unsustainable for a large number of cores. Point-to-point communication requires dedicated wires and specialized interfaces (which buses do not have) [102].

8.2 A Network on Chip: The Mesh

Networks on chips exploit parallelism to operate at lower non-Moore's law clock speeds but boost aggregate computation by performing such computations concurrently. While cores can always be added their effectiveness is only as good as the on-chip network that binds them. The network on chip paradigm is the current solution for the future of systems on chips. Networks on chips are scalable, reliable, and modular. Network on chip issues include chip area for the NOC, power consumption of the NOC, NOC latency, deadlock potential, and the choice of interconnection network [106].

To aid our discussion we will consider in detail the most popular and the most basic structured (or regular) interconnection network for on-chip networking: the mesh. A 3×3 mesh is shown in Fig. 8.2. It is basically a chessboard (Manhattan streets) pattern of interconnection. Intellectual property (IP) cores are connected to their north, south, east, and west neighbors through a network interface (NI). More specifically each network interface connects through a link to a nearby "router." Routers and associated mesh links form the actual network. It is most efficient for communications to take the shortest path in the mesh. Internally, each router consists of at most a 5×5 crossbar interconnection network (for up to four neighbors and a local connection), buffers, control logic, and registers (see Fig. 8.3).

The Network Interface breaks data into packets for transmission over the mesh and may also encode/decode packets in terms of error codes. The network interface may also break packets into smaller fixed length units called "flits" (FLow control digITs). A packet's flits consist of a header (or head) flit, a tail flit, and a number of

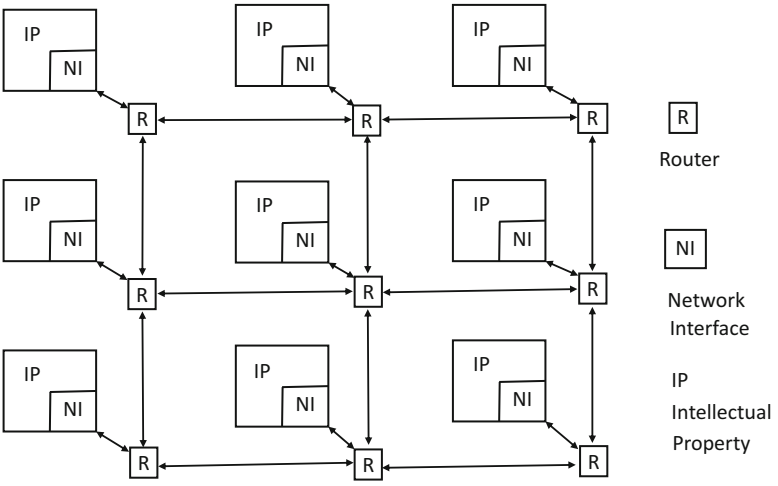


Fig. 8.2 A 2D 3 × 3 mesh network

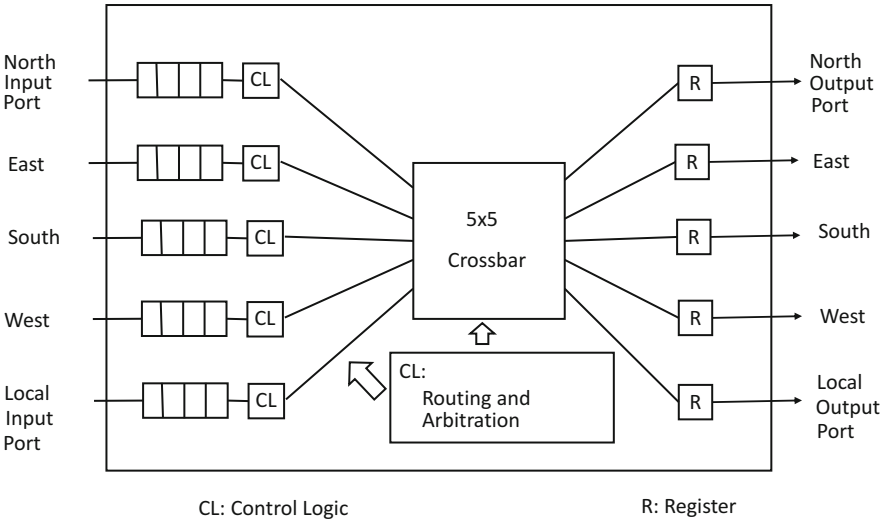


Fig. 8.3 Internal router structure for mesh node

body flits in between [30]. Flits entering a router are stored in buffers (see Fig. 8.3). Control logic routes each flit through the crossbar to enter a register and then be transmitted over an output link to either the next router along a path or to the network interface of the local IP core. The routing of flits is based on the (destination) header of the flits.

The control logic's routing decision can depend on the routing algorithm, arbitration rules, and downstream buffer availability. The diagram of Fig. 8.3 essentially

shows a crossbar with input buffering. Routing decisions are made at the input so flits are not backed up at the outputs of the router [30].

Large buffers can increase throughput and result in a lower delay but require a larger chip area and power usage [30].

Control logic functionality can be divided between routing and arbitration. The routing function generates the direction requests. The arbitration function creates a grant signal to indicate a winning request that can pass through the associated output direction (more than one flit/packet may wish to pass to an output at a time). Arbitration design must consider fairness to each packet/flit.

8.2.1 Switching Alternatives

Routing/switching of data can either be done on a packet basis or a flit basis. Switching on a packet basis can either be done using *store and forward* or *virtual cut-through* switching approaches. Under store and forward, the oldest packet switching approach, a packet must be completely received by a node before the node can begin to forward it to the next node along the packet's path. A more efficient alternative, in terms of latency (delay), is virtual cut-through switching. Virtual cut-through switching allows earlier parts of a received packet to be transmitted by a node before the node receives the complete packet from the previous node along the packet's path. It should be pointed out that this assumes that there is available buffer space at the next node, otherwise the packet is buffered, as in store and forward.

On a flit (FLow control digIT) basis, one can either use *wormhole routing* or a *virtual channel* approach [30, 106]. Under wormhole routing the flits are routed in a pipelined manner sequentially through the network. The first flit, which is a "header," reserves a channel in each router. The intermediate (or payload) flits follow on this path and the last tail flit frees the channel reservation. Thus the flits of a packet may be on different routers at the same time. In this sense wormhole routing is similar to virtual cut-through switching though it is done on a flit basis, not a packet basis. Wormhole routing allows compact and fast routers because it requires a smaller amount of buffer space than packet based approaches. However, head of the line blocking (HOL), which is well known in switching theory, can occur with wormhole routing since an (idle) packet may block a channel (at the buffer) even when other packets are available to make use of the channel.

A solution to this head of the line blocking problem is to use virtual channels [30, 106] (see Fig. 8.4). Under the virtual channel concept, n virtual channels are assigned to a physical channel, requiring n flit queues. These are in parallel (between the de-multiplexers and multiplexers prior to the crossbar in Fig. 8.4). Thus if a packet associated with a virtual channel is blocked, other flits from other virtual channels can be transmitted on the physical channel. Again, there is control logic in the virtual channel approach consisting of routing and arbitration functions.

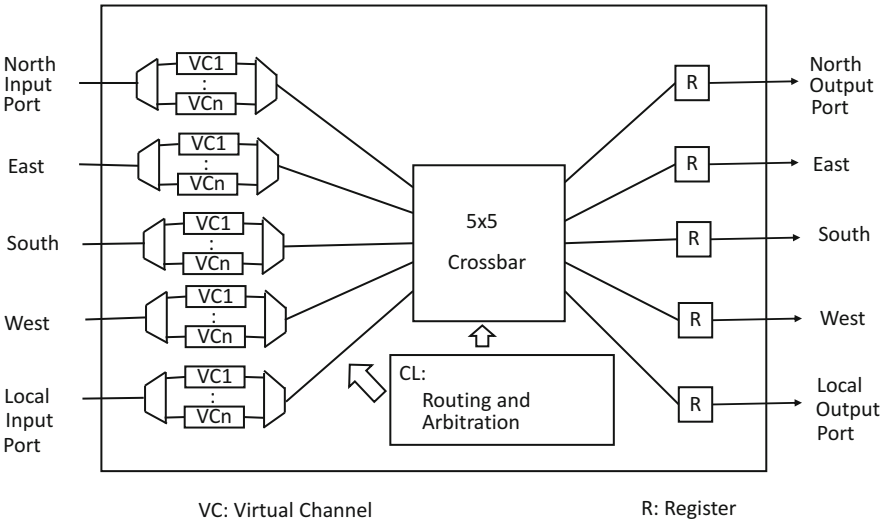


Fig. 8.4 Router with virtual channels internal structure. Here "CL" is control logic

8.3 Other NOC Interconnection Networks

8.3.1 Introduction

There are a large number of different types of interconnection networks that have been developed over decades. The applications for some of this early work were originally for telephone networks and later for multi-computer/multiprocessor networks. In this section we present some examples of these interconnection networks that have been used or suggested for NOCs [80, 150].

8.3.2 Mesh, Toroidal, and Related Networks

There are logical generalizations of the two dimensional mesh interconnection network. One generalization is simply to increase the number of dimensions. For instance, with interest in three dimensional integration circuits increasing, the three dimensional mesh network is the subject of some attention.

Another generalization of a two dimensional mesh is to introduce toroidal connections to create a toroidal network. Here there are direct horizontal links between left boundary nodes/cores and right boundary nodes. There are also direct vertical links between top boundary nodes and bottom boundary nodes. Thus each router requires a 5×5 crossbar (four connections to nearest neighbors and one local access connection to the local IP core). Toroidal networks get their name from the

Table 8.1 Some interconnect metrics

	Number of links	Node degree	Diameter
Complete interconnect	$N(N-1)/2$	$N-1$	1
Chain	$N-1$	$1-2$	$N-1$
Ring (1D torus)	N	2	About $N/2$
2D mesh	$2k(k-1)$	$2-4$	$2(k-1)$
2D torus	$2k^2$	4	$k-1$

Table 8.2 Some more interconnect metrics

	Avg. distance	Bisection width	Avg. distance $N = 1024$
Complete interconnect	1	$(N/2)^2$	1
Chain	About $N/3$	1	342
Ring (1D torus)	About $N/4$	2	256
2D mesh	$2k/3$	k	21
2D torus	$k/2$	$2k$	16

Table 8.3 d-dimensional meshes and torii

	Average distance	Node degree	Bisection width
d-dimensional mesh	$dk/3$	d to $2d$	k^{d-1}
d-dimensional torus	$dk/4$	$2d$	$2k^{d-1}$

fact they can be naturally embedded on the surface of a torus (i.e., donut shape). Multi-dimensional toroidal networks, including a three dimensional version, are also possible.

Tables 8.1, 8.2, and 8.3 show expressions for metrics for some interconnection networks [57, 126]. As a baseline in the table is a complete interconnection (where every pair of nodes has a direct link between them), a chain (i.e., linear network or 1 D mesh), and a ring (i.e., 1 D torus). Metrics that are sometimes used include:

- *Number of Links*: The number of (single hop) links.
- *Node Degree*: This is the number of connections of a node which may or may not include the local connection to a core.
- *Diameter*: This is the largest shortest path length between two nodes. Diameter is measured in hops (i.e., number of links traversed along a path).
- *Average Distance*: Average distance in number of hops between two randomly selected nodes.
- *Bisection Width*: The worst case amount of "bandwidth" between an equal division of the network into two parts.

In the tables N is the total number of nodes in an interconnection network, k is the size of a dimension (as in a k by k two dimensional mesh), and d is the dimension.

The presence of toroidal links lowers the average distance between nodes though the toroidal links' larger length (latency) may have to be taken into account. The metrics in these three tables are only for the interconnection network, not the local access (to the core) link. Note that the node degree in the tables could be increased by one to take into account the local connection to the nearest core. Similarly, other metrics such as number of links, diameter, and average distance could be modified.

8.3.3 Some Other Interconnection Networks

Another interconnection network is the fat tree. Here (see Fig. 8.5 for a binary fat tree) IP cores are placed at the leafs of a tree and the routers are placed at the upper nodes of the tree. The communication paths between IP cores climb up and down some parts of the tree. Particularly if paths between IP cores always utilize the tree root, more data carrying capacity is needed in the upper parts of the tree than the lower parts. Hence the name "fat tree" [76]. The thickness of the links in Fig. 8.5 indicates data rate.

Related to a basic fat tree, a butterfly fat tree is shown in Fig. 8.6. Note that in this butterfly fat tree three IP cores at a time are tied into a lower level router. Each lower level router is connected to two upper level routers. Thus each lower level router requires a 5×5 crossbar (three IP core connections and two upper level router connections). Each upper level router requires a 4×4 crossbar (four connections to lower level routers).

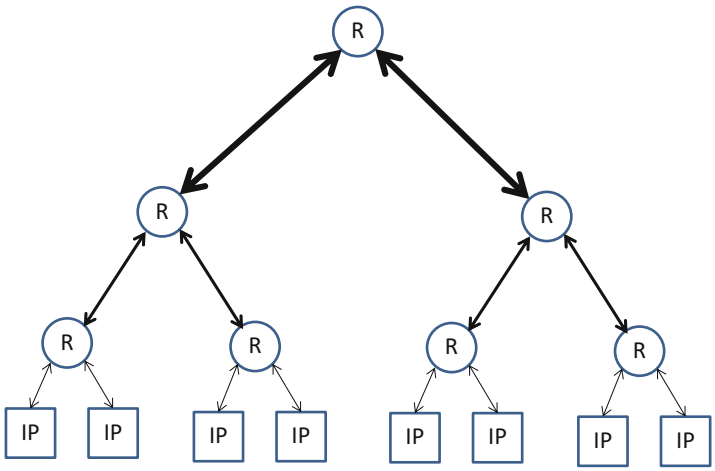


Fig. 8.5 Fat tree topology with routers (R) and intellectual property cores (IP)

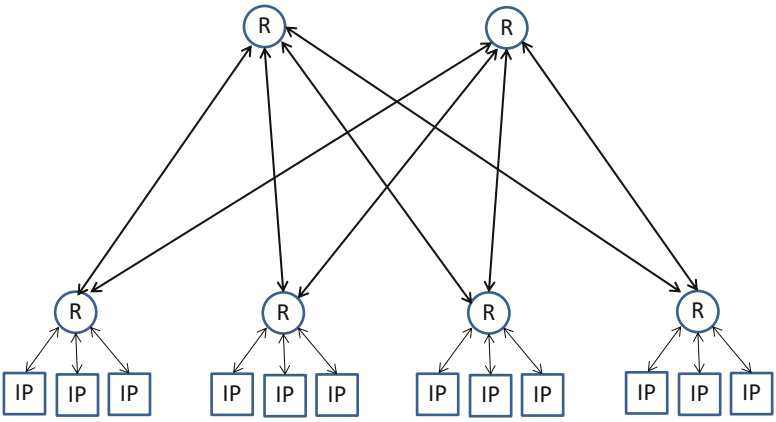


Fig. 8.6 Butterfly fat tree topology with routers (R) and intellectual property cores (IP)

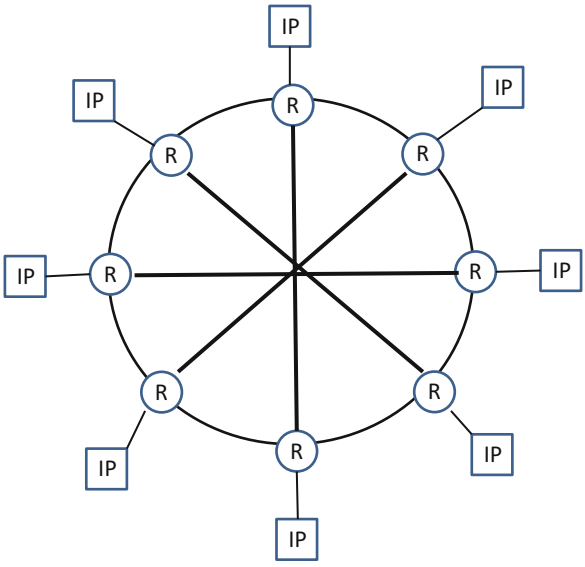


Fig. 8.7 Octagon topology with routers (R) and intellectual property cores (IP)

Finally another interconnection network that has been suggested for NOC use is the octagon topology (Fig. 8.7). There are 8 IP cores, 8 routers, and 12 two-way links. There are thus, at worse, two hops between any pair of routers. Each router requires a 4×4 crossbar (one connection to the local IP core and three connections to other routers). Simple algorithms can yield fast and efficient shortest path routing [150]. Octagons can be connected together to form larger networks.

Chapter 9
Space Networking

In the networks on chips chapter we looked at the smallest possible networks: networks on a single integrated circuit chip. In this chapter we examine the largest possible networks: space networks that span the solar system. We start more modestly though considering SpaceWire and SpaceFibre. This is internal networking technology for spacecraft. Designed in Europe, it is used by many space agencies and benefits from an elegant design. Then we consider space communications between earth and spacecraft. After a discussion of the background, we consider the Deep Space Network, Delay/Disruptive Tolerant Networks, the Bundle Protocol for space networking, and Contact Graph Routing. Let us begin!

9.1 SpaceWire

9.1.1 Background

During the early 1990s it became clear that there would be many advantages to a standardized onboard communication interface for spacecraft. It needed to support high speed (at the time 100 Mbps) data transfer between instruments and main memory. Some proprietary high speed networks were being created for the connection between instruments and main memory. In some instances MIL standard 1553 had been used for data collection for certain slower instruments but it generally wasn't fast enough for imaging and radar instruments.

Proprietary solutions had the usual disadvantages of this approach being expensive and lacking inter-operability. A standardized approach that would be used across many spacecraft had advantages in terms of cost, inter-operability, and reducing risk.

A possible solution was the developing IEEE 1355-1995 standard. This was a high speed data link using a special data-strobe encoding that was tractable

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_9

to realize in ASIC (Applications Specific Integrated Circuit) and FPGA (Field Programmable Gate Array) technology. However, the IEEE 1355-1995 standard had "contradictions, ambiguities and errors, some of which were serious" [111]. The University of Dundee in Scotland was contracted by the European Space Agency (ESA) to solve at least some of these problems. After analyzing IEEE 1355, a specification document for a spacecraft oriented network solution, which would become SpaceWire was written. SpaceWire was developed in consultation with European spacecraft engineers. In this section we follow the excellent treatment on SpaceWire in [111].

So what do we have? SpaceWire allows the creation of high speed/high performance onboard data handling systems. It provides data handling equipment compatibility and promotes the reuse of data handling equipment across multiple (and different) missions. Under SpaceWire, equipment is compatible at the subsystem and at the component levels. Equipment developed for a mission (processing units, memory, and telemetry for the downlink using SpaceWire interfaces) can be used on other missions. This maximizes the scientific work that can be done under a budget, improves reliability, and reduces risk.

9.1.2 SpaceWire in Detail

SpaceWire is specified in the European Cooperation for Space Standardization ECSS-E50-12A standard. SpaceWire is now used by ESA (European Space Agency), JAXA (Japanese Aerospace eXploration Agency), and NASA spacecraft and by others in the space industry.

SpaceWire used the DS-DE part of the IEEE 1355 standard as a starting point along with the TIA/EIA-644 Low Voltage Differential Signaling standard (LVDS). Also defined are appropriate cables and connectors for space applications.

SpaceWire uses high speed, bi-directional, full duplex data links operating at 2–200 Mbps.[1] The links interconnect equipment which have SpaceWire interfaces. Information is sent over links in a packetized manner using packets of any size. A packet is organized as:

<Destination Address><Cargo><EOP>

The destination address is a list of data characters that specify the network node(s) that the packet needs to go to. The destination address is either the identity code of the destination node or specifies the path that the packet should take in eventually arriving at the destination. The data to be carried in a packet is the "cargo." The EOP field indicates the end of the packet. See [111] for a discussion of packet addressing.

[1] Higher data rates are possible with the SpaceFibre extension discussed later in this chapter.

Networks can be built in many topologies with the use of point to point data links and routers. Timing and control information can be sent over SpaceWire. Timing codes can be sent with relatively low jitter.

There are a number of protocol layers to a SpaceWire network implementation [111]:

Packet Layer: Specifies how a packet is delivered from a source node to a destination node.
Exchange Layer: Characterizes the protocol for initializing a link, flow control, fault detection, and restarting a link.
Character Layer: Spells out the data and control characters used for the flow of data through a link.
Signal Layer: Specifies signal encoding, voltage levels, noise margins, and data speeds.
Physical Layer: Speaks of connectors, cables, and specifications for EMC.

As described in [111], the physical layer of SpaceWire is totally different from that of IEEE-1355, the exchange and signal layers between the two protocols are mostly different, and the packet and character layers have some major differences. SpaceWire is specifically designed for inter-operability with IEEE 1355 devices at the packet, exchange, and character layers.

The significant aspects of SpaceWire are [111]:

High Data Rate: SpaceWire can deliver 200 Mbps over 10 meters and go for greater distances at lower data speeds.
Links: Full duplex, bi-directional. The use of symmetric links makes flow control and building networks easier.
Implementation: Can be easily implemented in ASIC (application specific integrated circuit) or FPGA (field programmable gate array) technology.
Low Gate Count: Hardware only realizations can be implemented. A link interface needs between 5K and 10K gates.
Scalability: Links can be added to augment data carrying capacity.
Sharing Bandwidth: With multiple links between two endpoints, the total data rate of the links can be shared by all of the packets moving between the endpoints.
Fault Tolerance: With bandwidth sharing, if a link fails the remaining links will transport all of the packets moving between the two endpoints. This starts to happen immediately.
Topological Freedom: There is no constraint to network topology (buses, rings, trees, and hypercubes can be implemented, for instance). Links can be added to give more data carrying capacity or implement more fault tolerance.

9.1.2.1 Interfaces

SpaceWire utilizes two twisted pairs in each direction in its bi-directional links. As mentioned, Low Voltage Differential Signaling (LVDS) is used in the physical layer. Entire SpaceWire devices, which include LVDS drivers and receivers, can be

realized in a single chip. SpaceWire interfaces in general require roughly 5000–10,000 gates. A single radiation tolerant chip can hold one or more than one interface as well as a micro-computer or application logic [111].

Data-strobe encoding is used in SpaceWire. Here a serial data signal and strobe signal are sent on two differential pairs. The strobe is generated so that the clock can be recovered by simply sending the data and strobe lines through an exclusive or (XOR) gate. Without the need for a phase lock loop, SpaceWire interfaces are simply implemented as (digital) ASICs or FPGAs.

9.1.2.2 Router

Not too different from a Network on Chip (NOC) router (see previous chapter), a SpaceWire router consists of several SpaceWire link interfaces and a switch matrix (e.g., crossbar switch). Thus packets arriving at an input interface are directed through the switch matrix to an output interface. Naturally a (bi-directional) interface consists of an input and output port. A router checks the leading data character of the packet (a destination identifier character) to figure out the output of the router to which to send the packet.

Worm hole routing (see the previous chapter) is used in SpaceWire. This requires less memory in routers. This is an advantage for use in a radiation hardened chip.

9.1.2.3 Group Adaptive Routing

Routers in SpaceWire have an optional feature called Group Adaptive Routing (GAR). This provides both bandwidth sharing and fault tolerance.

Under GAR when two or more links are between two routers, the links can be configured through the routing tables as a "group." If a packet finds an output port of a group it wants to use is busy, any other output port of the same group can transport the packet.

Under the *bandwidth sharing* made possible by GAR, if there are N links in a group between two routers, data may be transported on any of the links (the links are shared). This leads to *N* times the carrying capacity of a single link between two routers.

For the *fault tolerance* feature, if any of N links in a group fails, data is automatically sent over the remaining links.

It should be noted that to implement bandwidth sharing one simply just configures any number of parallel links as a group in the routing tables. To implement the fault tolerance feature, one also simply lists a number of parallel links as a group in the routing tables.

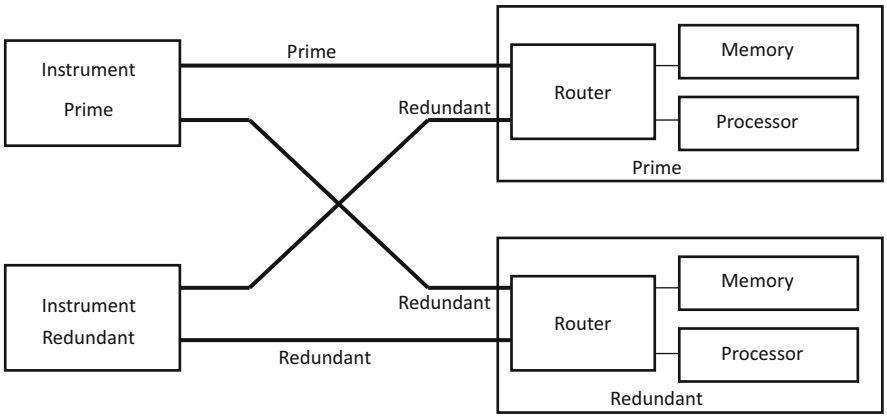


Fig. 9.1 Instruments and two data handling units

9.1.3 Some Configurations

In space missions redundancy is an important concern. A number of architectural configurations are discussed in detail in [111]. Two are mentioned here.

In Fig. 9.1, there is an instrument and a redundant instrument. There is also a prime data handling unit and a redundant data handling unit. Each data handling unit consists of router, memory, and processor. There is also a prime link and three redundant links implementing the connectivity redundancy as shown. The redundant instrument and memory are normally switched off to conserve power.

Including the router, memory, and processor in a single data handling unit reduces wire length and the mass of the SpaceWire links.

Following [111], some advantages of the architecture of Fig. 9.1 include:

- Supports multiplexing of several instruments.
- Supports a redundant instrument and a prime instrument.
- No single point of failure.
- Lower mass links.

Disadvantages include:

- Traffic should be examined to prevent temporary router blocking.
- Additional router power consumption.

Figure 9.2 illustrates other possible architectural features. There is a high data rate instrument connected to both prime and redundant data handling units. A new feature is the left router is concentrating the data flow of a number of (low to medium data rate) instruments (instruments 2 through 4). Concentrating this traffic with a single router reduces the size of the needed cable harness. Data can be sent from the left router to the data handling units over a single SpaceWire link.

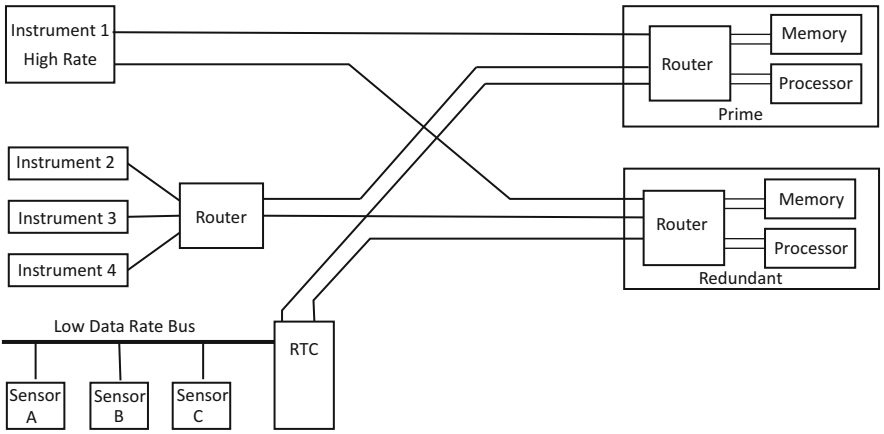


Fig. 9.2 More SpaceWire architectural features: data concentration

The pluses of this architecture and feature include:

- Smaller cable mass.
- There is a direct connection for the high data rate instrument(s) to the data handling units.
- Concentration for low to medium rate instruments.

Minuses include:

- There is no redundancy for the concentrating router.

The second new feature of the architecture of Fig. 9.2 is the concentration of low data rate sensor data on a bus through a Remote Terminal Computer (RTC). The remote terminal computer encapsulates data from the low data rate sensors into SpaceWire packets for transport to the data handling units.

Pluses to using the bus and RTC in this manner include:

- Legacy devices can be supported.
- Low rate sensor data is sent over SpaceWire to data handling unit. This longer connection is thus SpaceWire based.

Minuses include:

- Two types of networks.
- No redundancy for the bus or RTC.

9.2 SpaceFibre

9.2.1 Background

SpaceFibre, like SpaceWire, is a spacecraft onboard data link and network system also developed at the University of Dundee in Scotland for the European Space Agency. In this section we follow the extended discussion on SpaceFibre in [112, 113].

SpaceFibre is envisioned for use with high data rate instruments such as Synthetic Aperture Radar (SAR) and multi-spectral imaging instruments.

SpaceFibre operates at 2.5 Gbps[2] or about 12 times the capacity of a SpaceWire link. Data from a number of SpaceWire links may be concentrated onto a single SpaceFibre link. This thus uses a smaller cable harness mass. SpaceFibre can use either fiber optics or copper cables (for 5 m with copper and 100 m with fiber).

SpaceFibre can make possible a greater variety of spacecraft onboard communications applications using its internal QoS (Quality of Service) and FDIR (Fault Detection, Isolation, and Recovery) features and its backward compatibility with the popular SpaceWire.

9.2.2 SpaceFibre in More Detail

SpaceFibre has a number of innovations including its QoS mechanism which makes possible concurrent bandwidth reservation, priority, and scheduled QoS. This is an integrated QoS suite of features. The integration simplifies system engineering and testing and thus lowers costs.

Another innovation is integrated FDIR making possible galvanic isolation, transparent recovery from transient errors, and error containment in virtual channels and frames [113]. The onboard network robustness is improved through the built in FDIR and graceful degradation procedures implemented in network hardware. Thus FDIR software is streamlined which reduces development time and costs.

SpaceFibre also has low delay event signaling and it uses broadcast messages for time distribution.

All in all, a single SpaceFibre network can support multiple functions: high data rate payload data transport, transporting SpaceWire traffic, and deterministic delivery (useful for control signals, event signaling, and time distribution). At the packet level SpaceFibre is backwards compatible with SpaceWire. This allows simple interconnection of SpaceWire devices into a SpaceFibre network. This enables SpaceWire equipment to use the QoS and FDIR features of SpaceFibre.

[2]Future versions of SpaceFibre may operate at higher data rates.

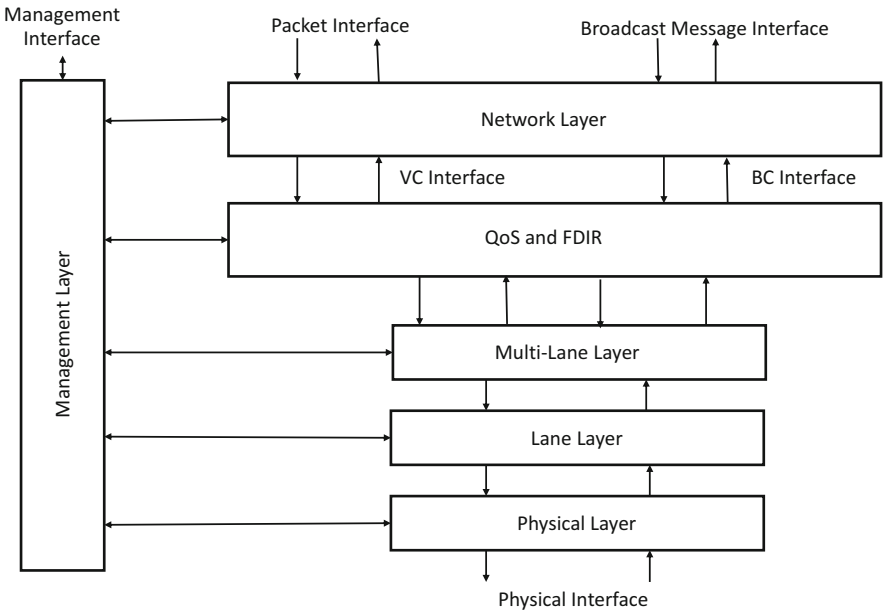


Fig. 9.3 SpaceFibre protocol stack

9.2.3 Protocol Stack

Figure 9.3 illustrates the SpaceFibre protocol stack. The Network layer supplies two services: Packet Transfer Service and Broadcast Message Service. SpaceFibre packets are carried over the SpaceFibre network using the Packet Transfer Service with the same packet format and routing techniques as with SpaceWire. Brief messages containing timing and synchronization information are broadcasted by the Broadcast Message Service to all network nodes.

Quality of service and flow control on SpaceFibre links are provided by the QoS and FDIR layer. It puts the data to be sent over a link into units called frames to provide QoS and also scrambles each packet to minimize electromagnetic emissions [113]. This layer also supports a retry feature: it retransmits frames and controls codes that don't arrive or arrive with errors.

The Multi-Lane layer supports multiple SpaceFibre lanes working concurrently (i.e., in parallel) to boost throughput. If a lane fails, this layer enables graceful degradation, using the remaining operating links. This is transparent to users.

Each lane is initialized by the Lane layer. The Lane layer also re-initializes a lane that has an error. Data is encoded by 8B/10B encoding which maps every 8 bits into 10 bits in such a way that the encoded stream has many transitions from 0 to 1 and 1 to 0 no matter the nature of the original data. This makes it easier for digital receivers to operate. The 8B/10B encoding is DC balanced so that AC coupling of SpaceFibre interfaces can be used.

The 8B/10B symbols are made serial in the Physical layer. At the receiver the Physical layer performs functions such as recovering the clock and data from the (serial) bit stream, finding symbol boundaries, and capturing the 8B/10B symbols. Again, SpaceFibre can be used with both fiber optics and electrical cables.

The management layer interfaces with all of the layers to enable configuration, control, and monitoring [112, 113].

As of 2015, parts of the SpaceFibre standard were either simulated, implemented, and reviewed or were in the process of being such. The ECSS (European Cooperation for Space Standardization) working group was to be convened to approve a finalized standard. The Network layer was to be a distinct standards document. Note that the Network layer has the same packet format as SpaceWire and has path and logical addressing features.

More extensive and excellent discussions of SpaceFibre including bandwidth reservation, IP core implementation, and SpaceFibre test equipment appear in [112, 113].

9.3 Space Communications

9.3.1 Background

Since the launch of Sputnik, there has been a need to communicate with spacecraft whether they be in some type of earth orbit or exploring planetary space. In some cases communications is the whole point of the spacecraft as in satellites providing communication connectivity, broadcasting to earth bound stations, or in bringing scientific data to earth from solar system exploration.

The early history of space communications involved radio contact to spacecraft as they came into view, telecommunications software and communication networks that were unique to each mission which led to costly system development and system operation. This motivated efforts at standardization, reuse of software and facilities, and unifying technologies and system architectures. These include efforts such as the Deep Space Network (DSN), Interplanetary Internet (IP), Delay/Disruption Tolerant Networking (DTN), and more [96].

As of 2013 the US National Aeronautics and Space Administration (NASA) operated three different networks to support various space missions [96]:

- Near Earth Network (NEN): For non-deep space missions using bands in the 2–8 GHz range.
- Space Network (SN): This is known as the Tracking and Data Relay Satellite System (TDRSS). It comprises seven geosynchronous satellites plus ground stations that use the 2, 13–15, and 26 GHz bands. The objective is to make available high speed satellite data rates [6 Mbps on the S band and 800 MBps on the higher Ka band (which is about 32 GHz)].

- Deep Space Network (DSN): Supports command, telemetry, and tracking services to a large variety of space missions. DSN services vary in their parameters. Command services utilize the S band (2–4 GHz) and X band (8–12 GHz) with Binary Phase Shift Keying (BPSK) modulation strategies. The size of a data unit is from 16 to 32,752 bits. The bit error rate for command services is a function of the signal to noise ratio at the spacecraft and is about 10^{-7} while service availability is between 95% and 98% of all times [96]. On the other hand, telemetry services in the S, X, and Ka bands of frequencies are used for both near-Earth and deep space missions. Modulation strategies include Phase Shift Keying (PSK), Binary Phase Shift Keying (BPSK), and Quadrature Phase Shift Keying (QPSK). The down link data rates have maximal values of 6 Mbps for deep space communication and 125 Mbps for near Earth communications. There is a minimum data rate of 10 bps. The frame (packet) rejection rate is around 10^{-4} to 10^{-5}. This determines the quality of the data returned. Service availability is the same as for command services [96].

There are standards and standard creating bodies to promote inter-operability among spacecraft and systems of the various space agencies. One standards body is the Consultative Committee for Space Data Systems (CCSDS). The CCSDS document library contains technical guidelines for data handling including data formats as well as specifications for the transport and physical layers. The CCSDS was founded in 1982. As of 2013 more than 500 missions have used standards from CCSDS. Also, the use of radio frequencies is standardized by the International Telecommunications Union (ITU), a UN agency. This includes global radio spectrum.

In terms of protocols, the Space Communications Protocol Specifications (SCPS) developed by CCSDS are new protocols and modifications to current protocols that yield better performance for Internet protocols used in space. These include (http://www.wikipedia.org) [96]:

- SCPS-FP: A space oriented extension to the file transfer protocol.
- SCPS-TP: Changes and extensions to the Transmission Control Protocol (TCP).
- SCPS-NP: Bit efficient network protocol.
- SCPS-SP: A security protocol similar to IPsec.

These protocols are used in different space environments. Some of these protocols are for near Earth situations and some are for deep space applications. They address critical difficulties in communicating through space: large propagation delays and high bit error rate (ber). There are other protocols that have been proposed and used as well (http://www.wikipedia.org) [96].

9.3.2 Deep Space Networks

In this section the focus is on the NASA Deep Space Network (DSN), which is used to provide communications to distant space missions. However, the similar

European Space Tracking (ESTRACK) network also has been successful and handled multiple deep space missions such as XMM-Newton and MARS Express. Other countries such as Russia, China, India, and Japan also have their own basic deep space infrastructure elements but do not have the volume of missions handled by NASA and ESA. Some of these missions are lunar missions, which are arguably not deep space missions.

The NASA Deep Space Network was created in the 1950s as a communications facility to support deep space missions. Deep space missions are visible for longer amounts of time than many near earth missions so few antennas are needed. However, because the deep space signals are so weak the antennas must be powerful and sensitive. The deep space DSN complexes are located approximately 120° apart. One is at Goldstone, California, one is near Madrid, Spain, and one is near Canberra, Australia. Each site has 8–14 h of viewing time. The Network Operations Control Team (NOCT) at the Jet Propulsion Laboratory's (JPL) Deep Space Operations Center controls the network.

Each facility is a distance from populated areas and located in semi-mountainous, bowl shaped sites to mitigate interference from civilization. Each site has at least four deep space transmitting and receiving stations with very sensitive receivers and large parabolic antennas that are high gain, steerable, and parabolic reflector based. High power uplinks use transmitters on 230 foot diameter antennas supplying 400 kW of power. Naturally, space mission transmitters are nowhere near this powerful, accounting for the weak signal received on earth (about 1000 billion weaker than a commercial TV signal received by a TV set) [96].

Microwave radio signals are used for communications by the Deep Space Network. Microwave signals travel in a straight line, resulting in less spread of the signal. Microwave signals can be focused by a lens or curved reflector to boost its brightness. The higher the frequency, the narrower the beam and the better the focus. There is much potential for interference to deep space microwave signals from other microwave uses (television, radio, cellular phones, and radar). Noise resistant telemetry coding, efficient antennas with large signal sensitivity, and low noise receivers are used to recover deep space signals in the crowded microwave radio environment.

As discussed in [96] the Deep Space Network has some limitations. Only the Australian facility is in the southern hemisphere which limits coverage. Almost three dozen spacecraft are tracked each year so the facilities must be shared. Moreover the DSN is expected to support twice as many missions in 2020 as it was in 2005. Older, legacy missions require support for many years and use of the larger antennas. Deferred maintenance of the 230 foot antennas result in them being out of service for several months at once. Replacements for the current dated antennas have been suggested in the form of array antennas [96].

9.3.3 Delay/Disruption Tolerant Networks

The term "Interplanetary Internet" was created by Vincent Cerf in 1997 to describe a future Internet providing connectivity throughout the solar system. The environment is characterized by frequent disconnection of links, and when they are connected links that are error prone and experience large delays. Links in space networks can be temporarily disconnected because of the interaction of the orbital motion of planetary bodies and spacecraft trajectory and location. Links can have high error rates which limit transmission speed. Round trip delay can be up to 40 min to Mars and more than 100 min to Jupiter. Also, each node in a space network may have different amounts of resources (such as memory) which affects transmissions.

Networking technology that can address these issues (both for space and terrestrial networks) goes by the name of Disruption or Delay Tolerant Networking (DTN) since early work by Kevin Fall, Scott Burleigh, and others [23, 42]. Depending on the different characteristics of different application areas, a DTN implementation is very much application specific [96]. Terrestrial applications such as wildlife tracking, underwater communications, military uses, and providing communication infrastructure to poor areas are also candidates for DTN.

Even within space networking, there are problems with different emphasis. For instance, Low Earth Orbit (LEO) satellites systems have satellites, as seen from earth, move across the sky, and lose connectivity to particular earth stations as they move below the horizon. They are thus "challenged" by disruptions to channel connectivity rather than extreme propagation delay (LEO satellites are close to earth). For a deep space mission propagation delay would play a bigger role.

9.3.3.1 Bundle Protocol

A major aspect of a solution architecture for DTN space networking is the use of a bundle protocol. The bundle layer, sitting above the transport layer, allows store and forward transmission of data and a custody transfer option to combat the difficult nature of the space environment. The store and forward feature deals with link interruption. Data transfer is allowed to be interrupted during disconnection periods and resumes when communication is reestablished. Nodes act as "custodians" of data under the custody transfer feature. Such nodes have the responsibility for guaranteeing the reliable delivery of data towards the destination. Thus no data is lost due to occultation or rotation in space [169].

The bundle protocol doesn't require temporal end-to-end connectivity as does the classic terrestrial Internet and is suited to space applications where connectivity is intermittent.

Figure 9.4 illustrates the use of the bundle protocol in a four node linear network. The bundle layer sits above the transport layer and below the application layer. The bundle layer has access to persistent memory.

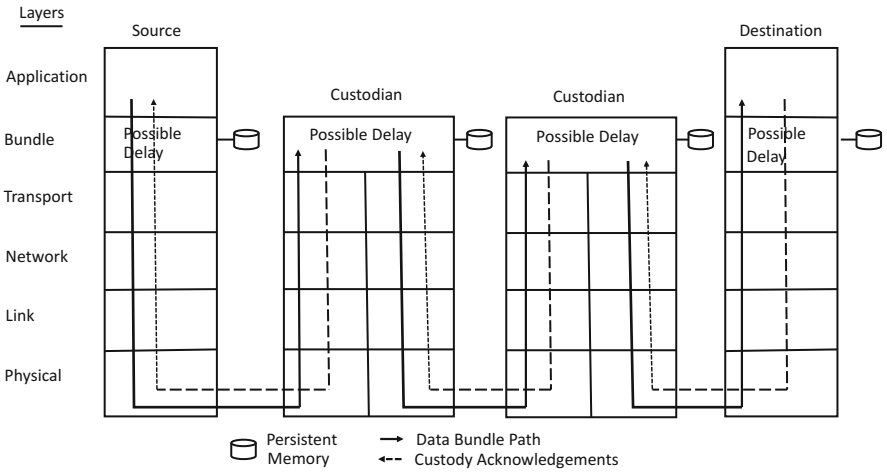


Fig. 9.4 Bundle protocol layered architecture

"Bundles" are variable length protocol data units (PDUs) carrying application data units. A bundle holds all the information that the application layer at the receiving end needs to use the data, without the need for additional data exchange (such as database queries or negotiation) with other nodes [169]. When activating custody transfer, custody transfer is enabled in the bundle layers of consecutive nodes called "custodians." This is initially at the request of the source application layer.

In terms of the operation of the bundle protocol in the context of Fig. 9.4, the application layer at the source node requests custody transfer and the source node is the initial custodian for a bundle. When it is possible, the bundle is sent to the second node which is the second potential custodian. Once the bundle is accepted by the second node a custody acknowledgement message is sent back to the source node which signals the source node that the bundle has been accepted by the second node. The source node can then discard the bundle from its persistent memory. If a retransmission timer ends before a custody acknowledgement is received by the source node, the source node transmits the bundle again [169].

Each consecutive pair of nodes along the path to the destination node repeats the procedure. Node i transmits the bundle to node $i+1$ and starts a retransmission timer. If the bundle is received and accepted at node $i + 1$, a custody acknowledgement message is sent back to node i which then discards its copy of the bundle. If the custody acknowledgement message doesn't arrive at node i by time the timer expires, node i then makes another attempt and retransmits the bundle.

Note that a custody acknowledgement is not an end-to-end acknowledgement. It simply indicates that the responsibility for end-to-end reliable delivery has been delegated to a new custodian (node). Responsibility for end-to-end delivery advances from node to node, from source to intermediate to destination nodes.

DTN bundles are stored at a node until they are delivered to the next node along the path to the destination or until application specified lifetime has expired [169].

A bundle node can be a thread on a system, an object in an object oriented programming environment, or a hardware unit [96].

9.3.3.2 Contact Graph Routing

In space, particularly deep space, Internet data needs to be routed through nodes without end-to-end connectivity. Rather, connectivity is intermittent but can be predicted through orbital mechanics calculations. Operating under a bundle protocol, nodes store data until "forwarding opportunities" [9, 46] (i.e., contacts) present themselves. Not all feasible contacts may be needed to route data. In the problem known as "contact plan design" (CPD), a contact plan (CP) is the optimized set of contacts that optimizes desired (performance) objective function(s) subject to some constraints. Early work sought to optimize (minimize) the number of contacts (i.e., connectivity) at the least cost. But beyond this one would like to consider capacity and fairness as additional optimization criteria as well as consider resource constraints (e.g., power consumption, transponder availability).

In Fig. 9.5 [45] four satellites (nodes) are shown in polar orbit. At time $k = 1$ N1 and N2 and separately N3 and N4 have connectivity. At time $k = 2$ only N2 and N3 have connectivity and at $k = 3$ we are back to the $k = 1$ connectivity situation. In this example it is assumed each satellite has dual antennas but because of power constraints a satellite has connectivity to only one other satellite at a time.

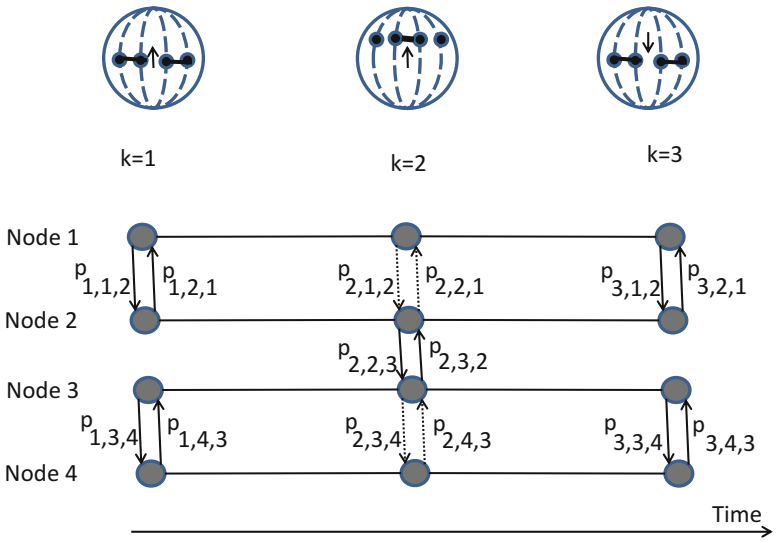


Fig. 9.5 Polar satellite model and finite state machine (FSM) representation

In Fig. 9.5 a finite state machine is also shown where each state is represented by a graph with arcs that indicate communication opportunities during a time interval. Here $p_{k,i,j}$ indicates the link between node i and node j at state (time period) k. If no contact is available, then $p_{k,i,j} = 0$. Also $p_{k,i,j} = a$ if connectivity between node i and node j is possible using interface a. Here, in the figure, solid lines are active links and the dotted lines are inactive links.

The contact topology can be represented by a three dimensional physical adjacency matrix. The contact plan can be put into this form by removing appropriate unused edges.

The topology can alternatively be specified by a contact list (see the Table 9.1 below).

For instance, here node 1 is connected to node 2 from time 0 to 1 ($k = 1$).

The contact list can also be drawn as a timing diagram as in Fig. 9.6. Here connectivity between node 1 and node 2, node 2 and node 3, and node 3 and node 4 is plotted as a function of time for the same polar orbit example. The subscripts of a contact "c" are source, destination, start and stop time.

Table 9.1 Contact list

Source–destination	Start–stop
1–2	0–1
3–4	0–1
2–3	1–2
1–2	2–3
3–4	2–3

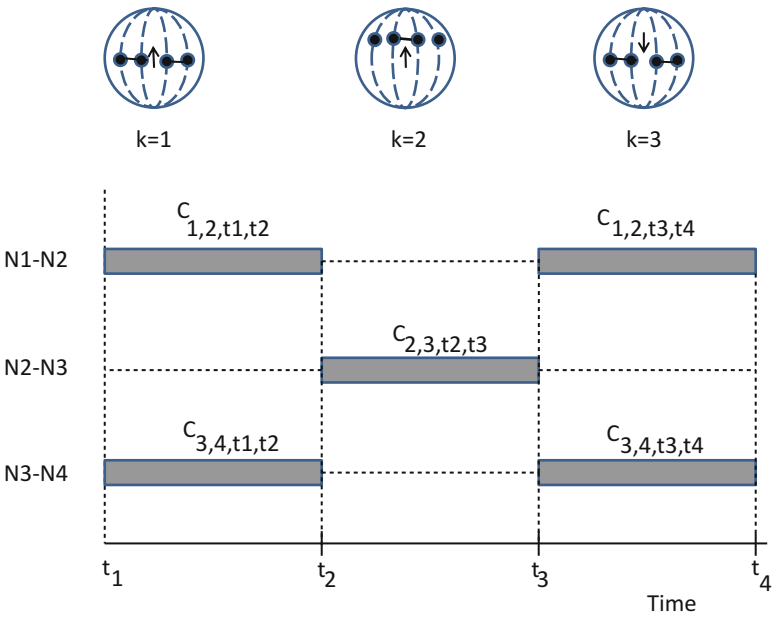


Fig. 9.6 Polar satellite model and contact timing diagram representation

Note that the contact list representation takes less space than the FSM adjacency matrix. However, the fine detail of the FSM can be easy to work with, particularly with Mixed Integer Linear Programs (MILP) optimization procedures [46]. It is simple to convert one representation to another though. In fact, the contact list format is used in the Interplanetary Overlay Network (ION) for contact plan storage and distribution [24].

9.3.3.3 Designing a Contact Plan

The feasibility of contacts in the future (the "contact plan") is based on communication system properties such as bit error rate, modulation, and transmission power as well as orbital dynamics such as position, range, and attitude (i.e., orientation of spacecraft and antennas) [46]. However, there are other constraints such as interference to and from spacecraft/sources as well as resource constraints such as node power or architectural restrictions. Also contact plans need to be distributed to nodes which is a subject of research [46].

Contact topology constraints can be divided into two types:

Time Zone Constraints (TZC): Time zone constraints are constraints that ban communication in a geographic area or time because of potential interference or agency based rules.

Concurrent Resource Constraints (CRC): This involves the quantity of concurrent communications (or contacts) a spacecraft is able to maintain at one time. For instance, a spacecraft may be able to use only one of multiple antennas at a time because of the power budget.

In contact plan design one generally evaluates the attractiveness of a number of feasible contact plans that satisfy the TZCs and CRCs in the contact topology space (see [46]). But with multiple feasible solutions one needs selection criteria [i.e., optimization criteria(ion)] to rank them and pick the best (optimal) solution. Possibilities include maximum contact time and contact assignment fairness, both of which depend on topological information. Routing or traffic information can also be considered. Mathematical programming or heuristic algorithms may be used to solve the contact plan design problem.

Chapter 10
Grids, Clouds, and Data Centers

10.1 Introduction

How does cloud computing differ from grid computing? Grid computing usually involves scheduled scientific or engineering computing done on a distributed network of computers and/or supercomputers. The applications for clouds not only are often a bit more mundane but also are more encompassing than grids. A basic idea is that organizations can rent time on an outside provider's data center(s) to host applications, software, and services that are more traditionally provided by in-house computing facilities. Data centers are locations housing thousands of servers and large amounts of memory. Data centers support grids and clouds. Some organizations such as Google, Facebook, and Microsoft run their own networks of data centers to provide their services. Both grids and cloud computing are accessed through the Internet.

10.2 Grids

10.2.1 Introduction

A grid is a distributed computing system that allows users to access large amounts of computer power and data storage over a network to run substantial computing tasks. Ian Foster, a leader in grid development, has written [47] that a grid must

- Provide resource coordination without a central control.
- Use standardized and open protocols and interfaces.
- Provide substantial amounts of service involving multiple resource types and non-trivial user needs.

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_10

As Schopf [130] points out, the idea of applying multiple distributed resources to work on a computational problem is an old one. It goes back at least to the mid-1960s and the "computer utility" paradigm of the MULTICS operating system and work on networked operating systems. Further work on distributed operating systems, heterogeneous computing, parallel distributed computing, and metacomputing further explored this area.

Work on the grid concept started in the mid-1990s. Among significant grid features that make it a distinctive problem area are:

- Grids allow site autonomy. Local sites have control over their own resources.
- Grids allow heterogeneity. Grid development work provides standardized interfaces to overcome site diversity.
- Grids are about data (representation, replication, storage in addition to the usual network and computer issues).
- The key person for grids is the user, not the resource owner. Earlier systems sought to boost utilization/throughput for the resource owner. In a grid, machines are selected to fulfill the user's requirements.

Grid computing provides advantages in terms of resource allocation, in particular for organizations spanning multiple time zones (in terms of utilization of off-peak resources) [91].

10.2.2 Grid Issues

A great deal of grid work is related to interfaces and software development to allow different sites to work in concert. However, the grid effort has taken time, enough time that some have questioned its practicality. An early discussion of the difficulties and challenges facing grid development appears in Schopf and Nitzberg. Here we mention some of these problems.

- In many cases, grids are being used to solve embarrassingly parallel applications rather than for coordinated distributed computing.
- Users often have to go through a great deal of work to achieve even basic functionality. It can be difficult for new user communities to move their community applications to a grid because to do this authentication mechanisms, job submission, and data access interfaces are not simple [20].
- Some science users see making applications suitable for the grid as a "distraction" from getting the science accomplished.
- Funding for adapting applications to a grid environment can be drained, if not blocked, by installation and administration problems.
- System administrators are used to having control over their local resources so grid software can be seen as "threatening."
- Setting up an account in a distributed grid can be complex.

An additional problem is that the number of users has not met the expectations of those who are in favor of grid technology [134].

Mattmann et al. [88] feel that the developers of grids have proposed domain specific software architectures for grids before having substantial experience in constructing such systems which has made the effort a risky and error prone venture.

However, the vision of successful grids may make overcoming these difficulties and growing pains worthwhile. Quite a few countries have invested in grid infrastructure to date. Efforts include the US National Science Foundation's National Technology Grid, NASA's Information Power Grid, Europe's Business Experiments in Grid (BEinGRID), the European Grid Infrastructure, World Communities Grid, and SETI@home. There are some grids for communities using large amounts of data such as, for particle physics, the LHC (Large Hadron Collider) Computer Grid. The Open Grid Service Architecture is an important grid standard announced in 2002.

10.2.3 Grid Architecture and More

The service oriented architecture of OGSA is illustrated in Fig. 10.1 In the figure, note that OGSA services are presented to the application layer. Also, OGSA services make use of web services. Note also this grid architecture has three hardware components: servers, storage, and networks.

Another way of describing the architecture of a grid is in terms of five layers [63, 88]:

1. Application: This top layer holds applications that use common services of the underlying grid.
2. Collective: This layer aggregates services provided by the resource layer. Information that is aggregated includes job status, resource monitoring statistics, as well as meta-data for each grid application.

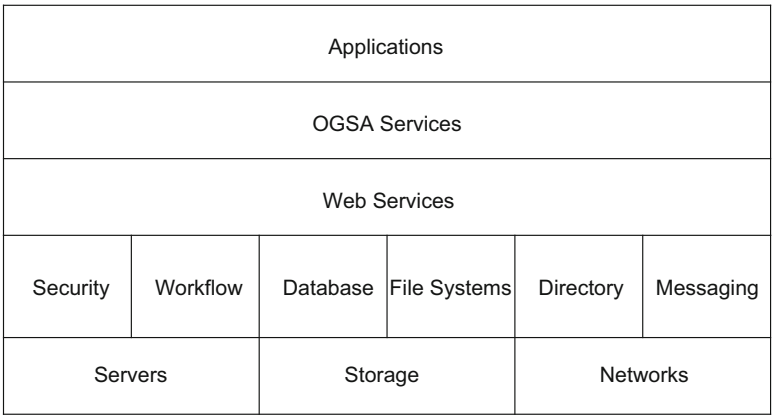


Fig. 10.1 Open Grid Service Architecture (OGSA)

3. Resource: This layer includes underlying computing resources (which are heterogeneous) and also includes a standardized interface for grid services communication.
4. Connectivity: Supplies communications, security and coordinates grid resource access to physical resources which appear in the fabric layer.
5. Fabric: This bottom layer includes items such as disk I/O, threading, and O/S resources which are available from grid nodes.

Some major types of grids are [88]:

- Computational Grid Systems: These systems solve large scale computational problems such as scientific work flow, climate modeling, earthquake prediction, and stellar simulations.
- Data Grid Systems: These systems normally collect, manage, and distribute large volumes of data/meta-data.
- Grid Monitoring Systems: These systems collect, analyze, log, and visualize monitoring data from grid resources.

Note that actual systems may do processing, deal with data, and include monitoring though the focus may be on one of these areas.

Grid middleware is a specific software product which enables resource sharing. Middleware is software that interfaces between an operating system (or database) and applications, particularly in a network context. Popular grid middleware are the Globus toolkit, gLite, and UNICORE (http://www.wikipedia.org).

Some argue economic model/policies for grids are useful for managing grid resources and managing supply/demand, improving the utilization of idle resources, and providing quality of service. There are many criteria that can be used to evaluate different economic models. These include such criteria as computation efficiency, evaluation of market price, the ability to handle large numbers of users, resource allocation efficiency, economic efficiency, and many others (see Haque [52] for a detailed listing and discussion).

Another important aspect of a grid system is a scheduler which matches resources to jobs over time. Scheduling is an optimization problem involving job requirements and job availability. Scheduling is branch of distributed computing and has a large technical literature associated with it [41].

A new extension to grid computing is mobile grid computing. This area merges the areas of grid computing and mobile computing. It involves using both data center computational resources and mobile computing resources such as tablets, laptops, and smart phones connected over a wireless network. New challenges in this area include networking, threats to the wireless network, quality of service, and monitoring (including security related monitoring) [147].

The main grid standards setting body is the Open Grid Forum. It was created from the merger of the Global Grid Forum and the Enterprise Grid Alliance in 2006. The Open Grid Forum is responsible for standards such as OGSA, Open Grid Services

Architecture (OGSA), Job Submission Description Language (JSDL), and GridFTP. Other organizations are also involved in the development of Grid standards.

For further information readers should see www.gridforum.org, www.globus.org, and [134].

10.3 Clouds

10.3.1 Introduction

The idea of cloud computing arose out of grid computing and may use a grid as its underlying network. In moving from grids to clouds the focus moves from an infrastructure that provides computing and storage resources to an infrastructure that is economic in nature and aims to deliver "more abstract" services and resources [91].

Again, under the usual cloud computing paradigm an organization can rent time on a provider's data center(s) resources to host applications, software, and services that to date have been provided on in-house computing facilities. There are three main types of cloud computing services [145, 172]:

Software as a Service (Saas): This allows a user (a consumer, student, employee of a business) to use applications accessed through the Internet that are run on the facilities of the entity supplying the service at a data center "inside the cloud." Thus the user only needs a web browser, not the actual application software. The user does not directly manage the cloud infrastructure (servers, storage, etc.).

Infrastructure as a Service (Iaas): Provides a complete computer infrastructure over the Internet (virtual computers, services, storage, etc.).

Platform as a Service (Paas): A user essentially has a "virtual platform" accessed through the Internet. The user can load applications using languages and/or tools that are supported by the cloud service provider. Users control applications but not the supporting infrastructure.

Other types of cloud computing can be defined. For instance, in terms of the Internet of Things and clouds one can define Sensing and Activation as a Service (SAaaS), Network as a service (NaaS), and Smart Objects as a Service (SOaaS) [26]. In general one can speak of XaaS (Anything as a Service) for cloud computing applications.

Zissis and Lekkas [172] describe four ways in which clouds can be used:

- Private Cloud: A private entity/organization uses cloud facilities dedicated for its own use. These facilities may be administered by the organization itself or a different entity and may be on site or off site.
- Community Cloud: The cloud facilities are shared by a community of several related entities. Again, it may be operated by the entities or a different entity and may be on site or off site.
- Public Cloud: Cloud facilities owned by an entity selling cloud services are offered to the general public, industry, government, or a selected group of such.

- Hybrid Cloud: An aggregation of at least two clouds of the previous types that support application and/or data portability for load balancing among the clouds.

Small and medium size cloud providers can organize themselves into larger cloud structures called inter-clouds, hybrid clouds, multi-clouds, and cloud federations [12].

10.3.2 Trade-Offs for Cloud Computing

Like any technology, cloud computing has a number of advantages and disadvantages [85, 158, 172]. Among the advantages are:

- Broad Access: This is made possible by the use of standard protocols.
- Cost and Economies of Scale: By "renting" services an organization can avoid large capital investments in infrastructure. Cloud service providers can take advantage of economies of scale by locating data centers on inexpensive real estate and in areas with low power costs and by aggregating the stochastic demand of a number of users into a centralized, expertly managed facility.
- Flexibility: Users can quickly have access to computing resources. Provisioning is elastic and fast. Services can be scaled according to client demand.
- New Services: New, innovative applications can be deployed in a straightforward manner. Such applications include parallel batch processing, business analytics, and social media such as Facebook and YouTube.

What of disadvantages? A major issue for organizations thinking about cloud computing is the loss of direct control of their data. Is the entity providing cloud computing services trustworthy? Do they run secure facilities? Are they liable to go out of business suddenly, causing a loss of data?

10.3.3 Cloud Principles

Computational principles for clouds include [127].

10.3.3.1 Multi-Tenancy

A cloud tenant is a user of cloud infrastructure and resources (IaaS). If multiple virtual machine owners use the same physical server/machine one has "multi-tenancy." A multi-tenant implementation benefits from the use of shared resources but gives the look of exclusive service to each tenant. That is, tenants need to be isolated from each other in terms of privacy, performance, and failure. Tenants must not be able to access each other's data/software. Also what happens to a tenant in terms of performance and failure should not affect other tenants [127].

10.3.3.2 Statistical Multiplexing

This is the idea that bursty work demands when aggregated create a less bursty and more consistent overall work demand. On average a service can meet many demands in an efficient manner because they are not all active at the same time.

10.3.3.3 Horizontal Scalability

Efficient execution of jobs in a cloud system requires efficient scaling across machines or "horizontal scalability." That is, at least some of a job(s) must run in parallel. One way of looking at a limitation of the ability of a cloud (or really any parallel system) to exploit parallelism is Amdahl's Law [8].

It can be expressed [127] as

$$T(n) = T(1)(\beta + ((1 - \beta)/n)) \tag{10.1}$$

$$\text{Speedup} = \frac{T(1)}{T(n)} = \frac{1}{\beta + \frac{1-\beta}{n}} \tag{10.2}$$

Speedup is the ratio of a job's execution time on one machine, $T(1)$, to the execution time on n (homogeneous) machines, $T(n)$. Here β is the fraction of a job that is serial (needs to run on one machine) and $1 - \beta$ is the fraction of a job that can run in parallel. So in the equations we have a weighted sum of the serial execution time on a machine and the parallel execution time on a machine. As an example, if $\beta = 0.10$, the maximum speedup no matter how many machines are used is ten (let n go to infinity).

Most programs have serial parts because distinct code sections may need to run serially to synchronize parallel computations. There are three possibilities with Amdahl's Law in terms of speedup:

Linear Scalability: This occurs if β is negligible. Using n machines results in an n times speedup.
Sub-linear Scalability: Here the speedup of a job grows more slowly than the number of processors used. This happens often because of overheads such as distribution overhead.
Super-linear Scalability: This surprising case occurs when the speedup grows at a rate faster than the number of processors used. It may occur, for instance, if there is a shared resource such as a shared cache that benefits from more usage [127]. It may also occur for some types of loads with a computational complexity that is non-linear in the size of the load [122].

In [127] there is an excellent discussion of the economic theory of cloud operations.

10.3.4 Cloud Monitoring

Monitoring clouds (or any large scale distributed system) is important for such functions as the design of clouds, troubleshooting, operation and maintenance, cost and traffic forecasting, security, testing and maintaining acceptable performance [158].

Cloud monitoring typically has three parts:

- Collection of data: The monitoring system either polls for data/state on machines or machines push data/state to the monitoring system.
- Analysis: This may involve graphing in simple cases or more complex and potentially computation intensive system analysis. Threshold analysis is often used (checking if a quantity exceeds or is below a pre-defined level). More intense computations may be challenging in their need for resources [158].
- Decision Making: Automated decision making is rare in monitoring systems which usually require a human in the decision loop. Autonomic decision making is an area of research [158].

Finally, among requirements for a cloud monitoring system are scalability, cloud awareness, fault tolerance, autonomic operation, the ability to handle multiple granularities, time sensitivity, and comprehensiveness. See [158] for a detailed discussion of these requirements and a discussion of actual monitoring systems.

10.3.5 Resource Provisioning

The resource provisioning problem for a distributed system such as a cloud is to allocate sufficient network and computational resources to allow efficient and effective operation at a reasonable cost.

In a cloud, resource provisioning involves selecting a cloud if there is a choice available, selecting a data center and selecting servers. The latter two problems in particular are provider problems. Algorithmic approaches to solving a resource provisioning problem include bin packing algorithms, greedy algorithms, graph theory based algorithms, virtual network embedding algorithms, and algorithms based on machine learning, control theory, or queueing theory. See [170] for a detailed discussion of these approaches.

10.3.6 Mobile Cloud Computing

Mobile cloud computing is an amalgam of mobile computing, cloud computing, and Internet technology. When integrated it provides a platform to users to provide information technology services at any location using the mobile Internet to a

cloud [69]. Such mobile cloud augmentation "employs resource-rich clouds to increase, enhance and optimize computing capabilities of mobile devices aiming at execution of resource-intensive mobile applications" [3]. Research issues include application development that is not simple and illicit access to data.

10.3.7 Cloud Reliability/Resiliency

Failures in cloud computing may be due to hardware or software, scheduling problems, server failure, power loss, denser system packaging, networking issues, cyber-attack, human errors, and correlated failures in either space or time. The reliability or resiliency of cloud computing systems can arise either in terms of security or in terms of resource/service failures [138]. With tens of thousands of processors in a cloud failures occur frequently and must be compensated for.

Reliability strategies include [33]:

- Failure prediction and resolution: This involves using system measurements to produce forecasts of failures in order to improve the preparedness of a cloud. Infrastructure providers may reorganize software components or the infrastructure itself to mitigate or eliminate failures a priori.
- Protection: The involves the use of redundant computation and communication resources to bring back services after a failure occurs. In distributed systems in general, not just for clouds, protection can be done in either of two ways, replication and checkpointing:

  (a) Replication: This is the most frequently used technique. It is done through either full or partial duplication of resources (machines, data, connections) to provide backup in case of failure. Some ways of doing this have certain efficiencies (such as RAID for disc memory).
  (b) Checkpointing: This involves saving the system state every so often in a different location and using it to start service again at that point if the original implementation fails.

- Restoration or recovery: This is a reactive effort to mitigate the effects of unspecified failures after they manifest themselves (which is different from protection where redundancy is deployed prior to a failure). For instance, one may reload a failed application or component.

Metrics for quantifying the degree of reliability involve survivability (the capacity of a system to survive a specific type of failure), recovery time (the time needed to recover), and cost (as in the cost of resources associated with a reliability technique) [33].

See [138] and [33] for more detailed discussions of cloud reliability and resiliency.

10.3.8 Cloud Security

Cloud security is a large topic. There are many issues and attacks that require attention. Security concerns include software, infrastructure, storage, and networking [142].

Attack types include tenant on tenant, provider on tenant, tenant on provider, and application level [10]. One can also speak of virtual machine to virtual machine (VM to VM) attacks, client to client attacks, and guest to guest attacks [58]. Attacks on storage may seek to obtain private data [64].

Security techniques to fight attacks include encryption, intrusion detection systems (IDS), signature techniques, intrusion prevention systems (IPS), access control, authentication, and trusted computing [10].

See these references for extensive discussions of cloud security.

10.4 Data Centers

10.4.1 Introduction

Data centers are networked collections of computers that provide the computational resources for web-hosting, e-commerce, grid, and cloud computing and social networking in a centralized location. Generic service platforms developed over time for this purpose include Sun's Grid Engine, Google's App engine, Amazon's EC2 platform, and Microsoft's Azure platform. A permanent data center may cover 300–4500 m^2 and be the home of thousands of servers. A percentage or two of US electrical demand comes from data centers [87]. In the following we largely follow the excellent treatment in [31, 61].

In terms of existing data center networks, Google is an example with a network of 36 production data centers (19 in the USA, 12 in Europe, 3 in Asia, 1 in Russia, and 1 in South America). They provide the infrastructure for Google offered services such as search, Gmail, and Google Maps. During 2016–2017 Google is building additional data centers in Oregon, Tokyo, and ten other locations [31].

Data centers can be built in dedicated buildings or as modular data centers (MDC) which utilize self-contained shipping containers (20–40 feet or so in length). A container includes servers, memory, networking equipment, racks, uninterruptible power supplies (UPS), a cooling system, and other data center components. Modular data centers can be deployed readily and support about six times more servers, in the same space, as a permanent data center. They have other advantages as well. On the other hand, modular data centers are costly, harder to maintain because of the cramped space and have some compatibility issues as time goes on [31]

A green data center uses energy efficient technology, green management, and renewable resources.

Virtualization, the ability of a data center to service many independent users while giving each user the impression of a dedicated facility, is important in data center technology for providing good server utilization and for making resource allocation flexible. Because of virtualization, data center management is not a simple problem. The expected trend in data center development is that data centers will become more virtualized, distributed, and a "multi-layer infrastructure" [61]. This will lead to a number of difficult technological problems.

10.4.2 Racks

The personal computer-like computers (servers) used in data centers are mounted in racks. A rack will actually contain not only the servers but also often storage and specialized devices. A standard rack is 78 in. high, 23–25 inches wide, and 26–30 in. deep. Assets mounted in a rack are measured in "U"'s (i.e., 45 mm or about 1.8 in.). A single or dual processor may be 1U. A four socket multiprocessor may be 2U or more. A standard rack can accommodate 42 1U assets [61].

Servers can be in a 13 inch high chassis. Six of these can be inserted into a rack since 6 times 13 is 78 in. (typical chassis height). A chassis will have a power supply, a backplane interconnect, fans, and management support. Sixteen 1U servers can be placed in a chassis so that a rack can hold 6 times 16 or 96 servers. "Blades" are modular assets in a chassis.

Racks vary greatly in their complexity. They may or may not include a metal enclosure, rack power distribution, air or liquid cooling at the rack level, a keyboard, video monitor, and mouse (e.g., kvm switch), and rack level management unit.

Power is a key consideration in data center design and operation. Racks may normally use 3–6 kW and blade server equipped racks' power usage may be substantially higher [131].

10.4.3 Networking Support

Data centers often are equipped with InfiniBand, Ethernet, lightweight transport protocols implemented over Ethernet, and/or PCI-Express based backplane interconnects. An issue is that with increasing transmission speeds, protocols need to be lightweight (i.e., simple). However, there is a need for some complexity because of security and other needed functionality.

There are at least four types of network access for a data center. This can lead to the use of several (even four) networking technologies in a data center [61].

1. Client–server network for data center access: This could utilize Ethernet or wireless LAN technology.
2. Server–server network: This operates at high speed and may utilize Ethernet, InfiniBand, or other networking technologies.

3. Access to storage: Historically this may use Fiber channel technology but storage access can also use technologies such as Ethernet or InfiniBand.
4. Management network: Could be Ethernet or a side channel on the main network.

Often the uplinks are "over-subscribed." One speaks of an over-subscription ratio as it may not be possible to achieve the full bisection bandwidth. Bisection bandwidth is the worst case bandwidth between a segmentation of a network into two equal parts. For instance, if 15 servers at 1 Gbps each share a 10 Gbps link, the over-subscription ratio is 1.5.

Data center hardware consists of switches (to interconnect elements), servers (including tower servers, rack servers, or blade servers), storage, racks, and cables (either copper or fiber).

Tree networks, traditionally used in data centers, have a problem in that they can have less cross section bandwidth than would be ideal. Fat trees and other types of interconnection networks have been suggested to alleviate this problem.

A classic data center architecture is shown in the figure [31]. It is a three layer, tree like, multi-rooted architecture where the uplinks of switches in the core layer provide connectivity to the Internet (Fig. 10.2).

The classic data center architecture has a number of issues including limited bandwidth, poor flexibility, low utilization, complex cabling, and high cost [31]. A number of proposals have been made to create better data center architecture. These include:

- Switch-Centric Architectures: These use switches improved in terms of networking and routing and also usually use unmodified servers. These can be tree like (such as fat trees and related proposals), flat switch-centric architectures (the typical three layer model is flattened to one or two layers), or unstructured switch-centric architectures (Jellyfish, for instance, logically creates a random graph for a flexible network size).

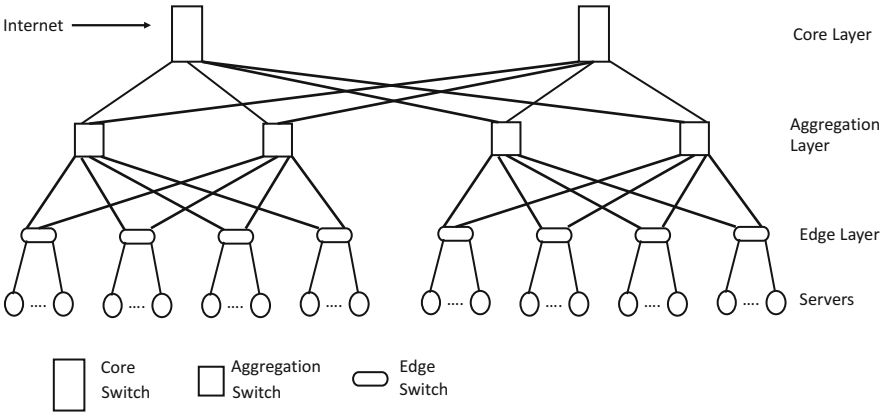


Fig. 10.2 Classic data center architecture

- Server-Centric Architectures: Here servers are in charge of networking and routing. Unmodified commodity switches just forward packets. These architectures take advantage of the high degree of programmability of servers compared to switches. Variants include server-centric architectures for mega data centers and server-centric architectures for modular data centers.

Future architectures may make use of optical or wireless connectivity within data centers. Among the benefits of optical networks are flexibility, higher bandwidth, and less power consumption and less heat than copper cables. On the downside there is about a 10 ms reconfiguration delay which is a problem for some applications. Electrical switches with optical transceivers are also costly.

Wireless connectivity in the 60 GHz band is being studied [31]. The data rate here can be as high as 7 Gbps.

See [31] for a detailed presentation of many data center architecture proposals as well as a tabular listing of many network metrics for switch-centric and server-centric architectures.

10.4.4 Storage

Storage capacity and data generation volume continues to increase at a great rate. Cisco believes that by 2019 each user will produce 1.6 Gbytes of consumer cloud storage traffic each month (in 2014 it was 992 Mbytes). Cisco also estimates that the amount of data created by Internet of Things devices will be 507.5 Zettabytes a year by 2019 (in 2014 it was 134.5 Zettabytes a year) [65]. "Zetta" is a prefix indicating a one followed by 21 zeroes.

Storage usually has used rotating magnetic technology to date. Because of rotating discs's mechanical construction sequential access is faster than random access. Storage can be responsible for 20–30% of a data center's power usage. A potential technology is solid state storage (non-volatile RAM or NVRAM). Kant [61] suggests that solid state storage will be in a "supporting" role for rotating discs.

In data centers one has [61]:

- DAS (Direct Access Storage): direct connection to server.
- SAN (Storage Area Network): storage (block oriented) exists across the network.
- NAS (Network Attached Storage): storage (files or objects) exists across the network.

Data intensive applications, streaming, and search create large loads on the storage system. It tends to be less expensive to carry all types of networking load on Ethernet but basic Ethernet does not have quality of service (QoS) support.

10.4.5 Electrical and Cooling Support

A medium sized data center can require several megawatts in peak power. Energy is consumed or wasted by servers, networking equipment, storage and cooling systems. Over the past few years a large effort has gone into minimizing energy usage by all of these systems including recovering energy from waste heat (see [124] for a discussion).

Power is provided to the data center on high voltage lines such as three phase 33 kV lines. Transformers on premises step it down to three phase power at 280–480 V. Power then goes to uninterruptible power supplies (UPS). The UPS output is often single phase power at 120–240 V. This is supplied to a power distribution unit which provides electric power to a rack mounted chassis or blade chassis. Power here is stepped down and converted to DC from AC to provide plus or minus 5 V or plus and minus 12 V. Voltage regulators on the motherboard change this to even lower voltages for the rails such as 1.1, 5, and 12 V.

With power conversion efficiencies of 85–95% at each step (and 50% or so at the motherboard rails) there is a great deal of energy loss and room for improvement.

Cooling can account for 25% or so of a data center's electric power usage. Air may enter racks at 21 degree centigrade and leave at 40 degree centigrade. Cooling necessitates building air conditioning units as well as large chiller plants, air re-circulation systems, and fans.

Improving power usage can be done in a number of ways [61]:

- Designing hardware and software for low power consumption.
- Designing the occurrence of hardware power states to minimize power consumption. This includes for CPU's, interconnection networks and for memory. An example might be a low power sleep state for a piece of hardware when not being actively used.
- In general, power appropriate use and regulation of data center infrastructure.

One metric for measuring the efficiency of a data center is power-usage effectiveness (PUE) [129]. This is the ratio of the total power used by a data center to the power consumed by just the computers and networking equipment. Chen reports the Facebook Pineville data center (powered by a solar array) has a PUE of 1.07, the Google Saint-Ghislain data center a PUE of 1.16, and the Microsoft Dublin data center a PUE of 1.25 [31].

10.4.6 Management Support

There is management controller within each server known as the baseband management controller (BMC). Functions of the BMC include [61]:

- Booting up and shutting down the server.
- Managing hardware and software alerts.

- Monitoring hardware sensors.
- Storing configuration data in devices and drivers.
- Remote management capabilities.

10.4.7 Security

As data centers increase in size, there is a need for scalable security solutions. One difficulty is that an attack may originate from an organization sharing a data center. Mitigating or preventing denial of service attacks is another concern for data centers [61]. See also the earlier section in this chapter on cloud security.

10.5 Conclusion

There is a synergy between grid, cloud, and data center technologies that bodes well for the future development of these fields.

Chapter 11
AES and Quantum Cryptography

11.1 Introduction

Network security is a very large technical area. In this section we look at two limited but important topics: the Advanced Encryption Standard (AES) and also quantum cryptography.

11.2 AES

11.2.1 Introduction

The Advanced Encryption Standard (AES) is the encryption standard approved in 2000 under the auspices of the US government, originally for civilian cryptographic use. In 2003 the United States government approved the use of AES for classified and secret information. The National Institute of Standards and Technology led the 3 year approval process for AES [34]. The Advanced Encryption Standard is incorporated in standards, algorithms, and requests for comments from IEEE (Institute of Electrical and Electronics Engineers), IETF (Internet Engineering Task Force), ISO (International Organization for Standardization), and 3GPP (Third Generation Partnership Project, see wireless chapter).

11.2.2 DES

The predecessor to AES was DES (Digital Encryption Standard). The National Bureau of Standards, which eventually became NIST, called for proposals for a block encryption standard (i.e., encrypting a block of data at a time rather than a

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8_11

continuous stream of data as a stream cipher would) in 1973 [25]. The only practical candidate was one from IBM. This was modified into what became known as DES.

There were some public issues concerning DES. The original 128 bit key was reduced to 56 bits in DES. Some changes were also made to scrambling boxes known as "S-boxes." Some felt that without knowing the reasons for the DES modifications it was difficult to make an assessment of its security. When differential cryptanalysis[1] was discovered some time later it was found that DES was more resistant to differential cryptanalysis because of the S-box changes. It was conjectured that IBM and NSA (the US code breaking and creation agency) knew of differential cryptanalysis and the changes were made for this reason [25].

As Burr [25] puts it, DES is the standard against which all block ciphers are compared.

In 2004 NIST withdrew DES through a version using three keys/steps of encryption and decryption, triple DES, was still approved because it was believed to offer better security than (single) DES.

11.2.3 Choosing AES

It is interesting to briefly discuss the selection process that led to the choice of the current AES (the summary here is largely based on the excellent discussion in [25]).

At the time of the selection of DES, the thought was that encryption is best done in hardware. But DES is not well suited for software operation. One needs to shuffle/scramble 4 or 6 bits. This is fast in hardware but slow on computers. Moreover triple DES is three times slower than DES. As the years went by, software encryption became more important.

NIST created a selection process for AES for federal and international business purposes. Among the properties an acceptable AES would need are [25]:

- Should be a block cipher.
- It should be at least as secure as triple DES.
- 128 bit block size.
- options for key sizes of 128, 192, and 256 bits.
- It should be unclassified and open to the public (not patented and royalty free).

Fifteen qualifying groups submitted proposals in 1998. The next year there were five final candidates:

- *MARS* from IBM (USA).
- *RC6* from RSA Data Systems (USA).
- *Rijndael* from Joan Daemen and Vincent Rijmen (Belgium).

[1] A technique where small changes to input are correlated with output changes in order to attempt to find the key.

- *Serpent* from Ross Anderson, Eli Biham, and Lars Knudsen (UK, Israel, and Denmark).
- *Twofish* from a team of American companies and academics (USA).

Burr's [25] article describes each contender.

In 2000 NIST selected Rijndael as the new AES. Rijndael was the most popular finalist in polls at the most recent AES conference. The international cryptographic community was well disposed towards it. Finally, the selection of a non-US cipher made international acceptance smoother. Rijndael was made the official AES in December 2001.

11.2.4 The AES Algorithm

The AES building blocks are shown in the accompanying figure [18, 137]. Key sizes may be 128, 192, or 256 bits. The number of round operations (repetitive loop) is 10, 12, or 14 for different key lengths (Fig. 11.1).

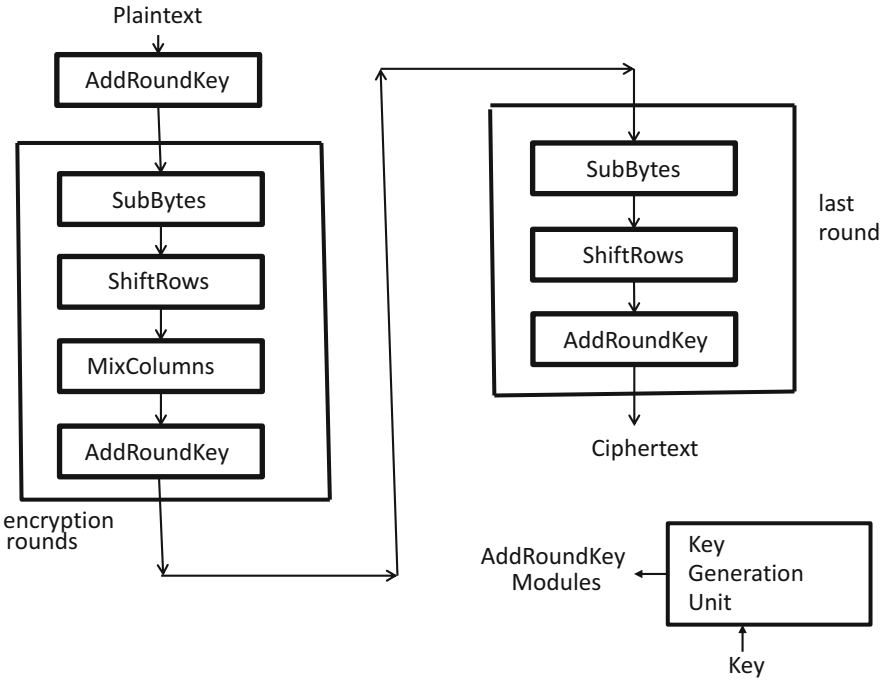


Fig. 11.1 AES encryption structure

Each round has four transformations [137]:

- SubBytes: Substitute bytes non-linearly. Processes each byte of the state independently using a substitution table (S-box). The S-box consists of two transformations: multiplicative inverse in the finite field GF(2^8) and an affine transformation.
- ShiftRows: A cyclic shift procedure in each row for four 4-byte data (offsets of 0 to 3).
- MixColumns: The 4-byte data blocks in each column are manipulated as coefficients of a 4-term polynomial and multiplies the data modulo $x^4 + 1$ with a specified polynomial.
- AddRoundKey: A bit-wise XOR operation on the data and round keys produced by the key generation unit from the secret key [137].

The encryption procedure of AES is run in reverse (inverted) to provide decryption functionality.

11.2.5 AES Issues

NIST's criteria for selecting AES included security, performance, and intellectual property.

11.2.5.1 Security Aspect

It was thought about 2003 [25] that an 80 bit key would provide sufficient protection against exhaustive search for a limited number of years. Moreover if Moore's law (computing power doubling every 18–24 months) remains true a 128 bit key would provide protection until 2066 and a 256 bit key would provide protection for several centuries. Of course what form technology will take centuries from now is hard to predict and breakthroughs like practical and robust quantum computing may change the situation radically.

Most symmetric key algorithms consist of multiple iterations/repetitions of a (scrambling) function called a round function [25]. People evaluating codes will try to find shortcuts on simpler versions of an encryption algorithm that use less than the full set of rounds. The attack will then be tried on the complete algorithm. People creating encryption algorithms estimate the number of rounds needed for security and add to these additional rounds as a safety margin.

An encryption algorithm is deemed secure if no shortcuts are known. The longer it has been evaluated the more confidence one has in its security. But encryption products are different from typical products. With a typical product it is designed, tested, and fielded. While surprises are certainly possible after it is fielded, they are rare. With an encryption algorithm no matter how much effort goes into its design, one is never quite sure in the back of one's mind if there is some clever way to crack it.

For the full versions of the final contenders in the AES competition no shortcuts were found. Attacks were found for versions of the final contenders with fewer rounds. Actually it wasn't possible to differentiate between the contenders on the basis of security. As reported by Burr [25] the best attack against Rijndael worked on 7 of 10 rounds and needed 2^{127} pairs, 2^{101} bytes of memory, and 2^{120} operations. Subsequent attacks aimed at breaking AES have only been successful on reduced versions of AES [34].

There are arguments about whether a simple or a complex encryption algorithm provides better security. A simpler algorithm can be better understood but some feel because it is simple a new attack may crack (break) it. For a complex algorithm if one part is broken the other parts may still provide backup security, some hold. But a complex algorithm may be difficult to fully analyze so there may be an unknown way an attacker may succeed in breaking it. The creators of Rijndael believe simplicity is the better virtue [34].

11.2.5.2 Performance Aspect

The performance of the candidate algorithms in the AES selection process were first discussed for 32 bit Pentiums. There was at least moderate performance for all of the contenders on Pentiums. All of the contenders had a better performance than triple DES. Performance was also examined based on RISC processors, Itaniums, 8 bit embedded microprocessors, digital signal processors (DSP), field programmable gate arrays (FPGA), and application specific integrated circuits (ASIC) [25].

11.2.5.3 Intellectual Property Aspect

It was desired that AES be available royalty free on world wide basis. This goal was made easier because DES patents had expired and there were quite a few algorithms available that were not patented. It was found by NIST that the final contenders' algorithms did not infringe on any existing patent.

There was also some question as to whether NIST should choose one winner or more than one. Straw polls at the third AES conference indicated that one winner was the community's preference. One reason [25] is that choosing several algorithms might lead to compatibility problems, require more chip real estate, and there would be a higher probability of intellectual property issues.

11.2.5.4 Some Other Aspects

It takes more than encryption algorithms to make secure systems. For instance, sending an email may require a public key transport algorithm, public key signature algorithm, a hash algorithm, and then the (AES) encryption algorithm [25]. Work has gone on in standardizing other strong algorithms in addition to encryption algorithms.

Another issue is that block ciphers such as DES and AES can be used in a number of well-known "modes" of operation. These include [25, 68]:

- *Electronic Code Book Mode*: A plaintext number of bits is mapped into the same number of encrypted bits. A weakness is that an attacker may swap sections of the code with code of their own.
- *Cipher Block Chaining Mode*: One linearly chains block ciphers so replacing a block causes unintelligible code after that point.
- *Cipher Feedback Mode:* Uses feedback in the coding process for byte by byte encryption.
- *Stream Cipher Mode:* A simple way to convert any block cipher to a stream cipher (i.e., encrypting a continuous stream of data rather than a block at a time).

A weakness of chained and feedback modes is that they can't be parallelized and they can't be arbitrarily accessed (one needs to start at the beginning). To address this a counter mode was introduced which is parallelizable [25].

Other modes have been proposed to NIST. Some are parallelizable and do encryption, authentication, and integrity protection for just a bit more than the cost of encryption. However, the inventors filed patents on these so there are some intellectual property issues.

11.3 Quantum Cryptography

11.3.1 Introduction

Quantum cryptography uses principles from quantum mechanics to assure security. Traditional cryptographic systems assure security by making a brute force attack (i.e., trying all possible keys) take so long as to be impractical. This is usually done by relating the encryption to well-known mathematical problems that are known, as long as people have worked on them, not to have a simple solution. Quantum cryptography uses physical quantum principles to prevent easy unauthorized decryption.

It should be pointed out that although current traditional cryptographic algorithms are not susceptible to cracking by today's computers, the situation may be different decades from now. Computers become faster year by year and over decades they will be significantly faster. This is a problem because information encrypted today often needs to be secure for decades. Moreover, if fully capable quantum computers (computers based on quantum computation) are developed over the next several decades they will easily crack certain (but not all) of today's ciphers. This leads to post-quantum encryption (see section below). Post-quantum cryptography involves encryption algorithms being developed today to resist unauthorized decryption in a future world of powerful quantum computers.

11.3.2 Quantum Physics

Quantum mechanics (or quantum physics) involves the behavior of small discrete amounts of energy or matter (e.g., photons and atoms). It was the subject of much research in the last century and this century. Work started with research by Max Planck on black body radiation in 1900 and by Albert Einstein on the photo-electric effect in 1905.

In standard digital logic systems (the building blocks of today's computers) information is represented by the two logic values of "0" and "1". In quantum information processing information is represented by "qubits." Unlike the binary nature of digital logic bits, qubits exist in a simultaneous combination of states. This is often discussed in terms of a two dimensional complex vector space.

When one attempts to measure a qubit it collapses into one of two states.

This all has some implications. Quantum computers are powerful in principle because they can sort through many states (solutions) concurrently. On the other hand, the measurement of a quantum state "disturbs" the state, so perfect copying is not possible (e.g., if using photons and optical fiber straightforward repeaters can't be used) [81].

11.3.3 Quantum Communication

Quantum principles can be used for quantum communication in the following ways [56]:

1. Quantum key distribution (QKD) to provide secure distribution of keys to be used in conventional cryptographic systems. This will be discussed in the next section.
2. Quantum teleportation for moving information in quantum computers. Quantum teleportation is a means of sharing quantum information (such as the state of a photon or atom) between two locations. This requires a classical information link so information transfer does not occur at a faster than light speed. In spite of the name, quantum teleportation refers to a transfer of information (not matter as in the Star Trek television series).
3. Entanglement swapping for setting up long distance quantum networks. "Entanglement" is a condition where the properties (position, momentum, spin, and polarization) of particles at a large distance are correlated. Measurements made on one particle of an entangled pair of particles are known by the other particle of the entangled pair (http://www.wikipedia.org). Quantum repeaters can't be built because the state of quantum particles can't be copied. But particles on either end of a network can be entangled (using a series of entangled network segments between them) to achieve a repeater effect and thus lay the foundation of a quantum network.

4. Quantum "seals" for maintaining the integrity of physical boundaries. Using quantum principles, quantum seals allow the testing of the soundness, authenticity, and physical layer security of a communication channel.

In the next section, quantum key distribution, the most developed of these techniques, will be discussed.

11.3.4 Quantum Key Distribution

Quantum key distribution (QKD) is a quantum cryptographic technique used to distribute keys for conventional cryptographic systems. But why are keys important? The whole basis of the security of classical encryption systems is in the key. For the older symmetric key model the same key is used for both encryption and decryption. In the more recent public key cryptographic systems there are two keys: often a public one for encryption and a private one for decryption.

It is often assumed that the method of encryption/decryption is known to attackers and ciphertext (the encrypted text) is also known to attackers. The security of the cryptographic system is in the key(s). But how does one distribute the keys to users in a secure manner? This is the well-known key distribution problem. There are five basic approaches to key distribution [7].

- Classical information theoretically secure key agreement methods.
- Classical computationally secure public key cryptography.
- Classical computationally secure symmetric key cryptography.
- Quantum key distribution (QKD).
- Trusted couriers.

With the large volume of data that compact storage devices can hold the use of trusted couriers is not beyond the realm of possibility. However, in this section the emphasis is on QKD. Actual QKD systems have been in use for a number of years.

The idea for QKD can be traced back to work by Stephen Wiesner about 1969 on unforgeable bank notes. The first cryptographic application of this, QKD, was published by C.H. Bennett and G. Brassard in 1984 and the work is known as BB84.

The original BB84 paper used photon polarization states to transmit information but any two pairs of conjugate states can be used. The sender (called Alice in the cryptographic literature) transmits quantum states using photons to Bob (the receiver) over a quantum communication channel. Let's think of this channel as implemented on a fiber but it could be implemented on a free space link.

As an example, one can use a rectilinear basis [i.e., photons polarized or made directional as vertical (0°) or horizontal (90°)] and also use a diagonal basis of 45° and 135°. In the tables in our example the use of a rectilinear basis is indicated by $+$ and the use of a diagonal basis is indicated by $\times$.

In our example Alice and Bob use a series of rectilinear and diagonal filters. If a vertically/horizontally polarized signal is applied to a rectilinear filter the

Table 11.1 Bases

Basis	0	1
+	↑	→
×	↗	↖

Table 11.2 A QKD example

Alice's bits	1	0	1	0	0	0	1	1
Alice's random basis	×	×	+	×	×	+	+	×
Alice's transmission	↖	↗	→	↗	↗	↑	→	↖
Bob's basis	×	+	+	×	×	+	×	+
Bob's measurements	↖	↑	→	↗	↗	↑	↖	↑
Basis exchange								
Shared secret key	1		1	0	0	0		

polarization (i.e., angle) is preserved when the photon leaves the filter. The same is true when a diagonally polarized photon is applied to a diagonal filter: the angle of the diagonally polarized photon is preserved by the filter. On the other hand, if a diagonally polarized photon is applied to a rectilinear filter, then there is a 50% chance a vertically polarized photon is produced by the filter and a 50% chance a horizontally polarized photon is produced by the filter. Likewise if a rectilinearly polarized photon is applied to a diagonal filter, then there is a 50% chance the filter produces a 45° photon and a 50% chance of producing a 135° photon.

For the example to be presented, Table 11.1 shows the two different bases and the encoding.

So, for instance, under a rectilinear basis, a "0" is represented by a vertically polarized photon. Likewise, under a diagonal basis a "0" is represented by a photon polarized at 45° and a "1" by a photon polarized at 135°.

In our example (Table 11.2) Alice and Bob want to establish a shared secret key. Alice generates a random series of bits. She also randomly picks a basis (rectilinear or diagonal) to use for each bit. Now using Tables 11.1 and 11.2, in the first column of Table 11.2, to transmit a "1" with diagonal polarization, Alice transmits a diagonally polarized signal at 135°.

Bob receives the photons with his own random series of filters. He doesn't yet know Alice's choice of filters so some of his guesses will match Alice's and some won't.

In the first column of Table 11.2 there is a match (both choose diagonal filters) so Bob measures correctly a photon which is diagonally polarized at 135° (a "1"). In the second column Bob guesses the wrong basis so he measures (with 50% probability) a vertically polarized photon.

At this point Alice and Bob exchange the bases they used over a non-secure channel. They only keep the corresponding bits where they agreed on the choice of filter. This is their shared secret key.

In the cryptographic literature "Eve" is a person trying to illegitimately acquire the message. Even if she had access to the transmission of polarized photons between Alice and Bob she can't read them because measuring them will change them and Bob and Alice could detect this. She doesn't have access to Bob's measurements so in theory she doesn't know the secret key.

There is also a way for Alice and Bob to check for the presence of an eavesdropper. They compare a specific subset of the key string. An eavesdropper in gaining information on the photons' polarizations will cause errors in the measurements made by Bob. However, such errors can also be produced by noise in the transmission line and by detector imperfections.

If the percentage of errors is below a certain threshold, two procedures can be used to remove the errored bits and make Eve's knowledge of the key to be an arbitrarily small value. These procedures are reconciliation and privacy amplification [109]:

- Reconciliation: This procedure removes errors due to bad choices of measurement basis, errors due to channel noise, and errors due to eavesdropping. This procedure is a recursive search for errors. Alice's and Bob's key sequences are divided into blocks and the parity of blocks is compared. When parities do not match blocks are divided into smaller blocks in a binary search approach to find errors. A cascade approach (http://www.wikipedia.org) is used to find multiple errors in the same block.
- Privacy Amplification: Reconciliation is done over a classical (insecure) channel. Thus an eavesdropper may gain certain information on the key by monitoring this channel as well as the quantum channel (though this later monitoring will introduce errors). Under privacy amplification Alice's and Bob's keys are used to produce a shorter key in a manner so that Eve has arbitrarily small knowledge about the key. This is done using a universal hash function (http://www.wikipedia.org) which essentially maps the original keys into a shorter key with a known probability of the amount of Eve's knowledge.

11.3.4.1 Implementation

Quantum key distribution is the most developed application of quantum cryptography to date. A number of companies offer QKD systems. A number of systems have been fielded such as the DARPA Quantum Network which is a 10 node network in Massachusetts. There are also systems in Switzerland, Tokyo, and Los Alamos. Systems have been fielded with key bit rates on the order of 10^4–10^6 bps (higher bit rates for shorter distances).

11.3.4.2 Absolute Security?

It should be noted that the security of a QKD system can be subverted due to threats such as hacking QKD computers and threats peculiar to quantum systems. Scarni and Kurtsiefer [128] feel that these threats can be protected against but claims of absolute security have to be put into this context.

11.3.5 Post-Quantum Cryptography

Much information that is encrypted should remain secure for decades. The development of capable quantum computers would make it possible to decrypt certain currently used encryption algorithms by using techniques such as Shor's algorithm [139]. Capable quantum computers may not developed until 2050 [82] but this would make information encoded decades earlier prematurely accessible.

Therefore research is underway on cryptographic algorithms that can be used today that would be resistant to the eventual use of quantum computers. This is post-quantum cryptography. For instance, the US National Security Agency announced in 2016 a "transition to quantum-resistant algorithms" for one of its cryptographic suites.

There is a misunderstanding that quantum computers can solve currently intractable NP complete problems in polynomial time. This is believed, or at least suspected, to be false (http://www.wikipedia.org). Thus it is possible to develop quantum computer resistant encryption algorithms today. Possible algorithm types include [82, 83]:

- Code based (error correction): Based on matrix multiplication and vectors, these use error correcting codes to produce public keys from private keys with errors injected into matrices on purpose. However, such public keys consist of millions of bits. A version of this approach was recommended by Europe's Post-Quantum Crypto Project [82].
- Hash based: Hash schemes, developed in the 1970s, are often used for message authentication (guaranteeing a transmitted message has not been altered). They can also be used to create public key cryptographic systems. However, keys are large and in a process called "chaining" every time a message is sent the public key must be recomputed and included in the next message. This adds overhead to implementations.
- Lattice based: These are based on the mathematical problem of finding shortest vectors (closest vector to a point) in a lattice consisting of points in n-dimensional Euclidean space. There are trade-offs between different types of lattice encryption algorithms and the sensitivity of some algorithms needs to be better established.

- Multivariate quadratic equation: These are based on the NP complete/NP hard nature of solving quadratic systems. Many types of encryption schemes can be developed on this basis.

In particular, lattice based and multivariate quadratic equation encryption algorithms assume their respective mathematical problems will remain difficult to solve into the future.

11.4 Conclusion

For a very long time cryptography has benefitted from the initiative shown by both attackers and defenders. This is likely to continue into the future.

References

1. 4G LTE advanced - what you need to know about LTE-A. 4G.co.uk, December 16, 2015
2. H.S. Abbas, M.A. Gregory, The next generation of passive optical networks: a review. J. Netw. Comput. Appl. **67**, 53–74 (2016)
3. S. Abolfazli, Z. Sanaei et al., Cloud-based augmentation for mobile devices: motivation, taxonomies and open challenges. IEEE Commun. Surv. Tutorials **16**(1), 337–368 (First Quarter 2014)
4. S. Ahmadi, An overview of next-generation mobile WiMAX technology. IEEE Commun. Mag. **47**, 84–98 (2009)
5. I.F. Akyildiz, D.M. Gutierrez-Estevez, E.C. Reyes, The evolution to 4G cellular systems: LTE advanced. Phys. Commun. **3**, 217–244 (2010)
6. I.F. Akyildiz, D.M. Gutierrez-Estevez et al., LTE-advanced and the evolution to beyond 4G (B4G) systems. Phys. Commun. **10**, 31–62 (2014)
7. R. Alléaume, C. Brassard et al., Using quantum key distribution for cryptographic purposes: a survey. Theor. Comput. Sci. **560**, 62–81 (2014)
8. G.M. Amdahl, Validity of the single processor approach to achieving large scale computing capabilities, in *Proceeding of the Spring Joint Computer Conference* (ACM, New York, NY, 1967), pp. 483–485
9. G. Araniti, N. Berzirgiannidis, E. Birrane et al., Contact graph routing in DTN space networks: overview, enhancements and performance. IEEE Commun. Mag. **53**, 38–46 (2015)
10. C.A. Ardagna, R. Asal et al., From security to assurance in the cloud: a survey. ACM Comput. Surv. **48**(1), 1–50 (2015). Article 2
11. G. Armitage, MPLS: the magic behind the myths. IEEE Commun. Mag. **38**(1), 124–131 (2000)
12. M.R.M. Assis, L.F. Bittencourt, A survey on cloud federation architectures: identifying functional and non-functional properties. J. Netw. Comput. Appl. **72**, 51–71 (2016)
13. B.N. Astuto, M. Mendonca, X.N. Nguyen et al., A survey of software-defined networking: past, present, and future of programmable networks. IEEE Commun. Surv. Tutorials **16**(3), 1617–1634 (Third Quarter 2014)
14. A. Bacioccola, C. Cicconetti, C. Eklund, L. Lenzini, Z. Li, E. Mingozzi, IEEE 802.16: history, status and future trends. Comput. Commun. **33**, 113–123 (2010)
15. M. Baker, A. Apon, C. Feiner, J. Brown, Emerging grid standards. Computer **38**(4), 43–50 (2005)
16. O. Banimelhem, J.W. Atwood, A. Agarwal, Resiliency issues issues in MPLS networks, in *Canadian Conference on Electrical and Computer Engineering, CCGEI 2003*, Montreal, QC, May 2003, pp. 1039–1042

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8

17. D. Barak, Verbs programming tutorial. Open SHMEM 2014, Mellanox, 2014
18. S. Baskaran, P. Rajalakshmi, Hardware-software co-design of AES on FPGA, in *Proceedings of ICACCI'12*, August 2012, pp. 1118–1122
19. B. Bellalta, L. Bononi, R. Bruno, A. Kassler, Next generation IEEE 802.11 wireless local area networks: current status, future directions and open challenges. Comput. Commun. **75**, 1–25 (2016)
20. M. Bencivenni, D. Michelotto, R. Alfieri et al., Accessing grid and cloud services through a scientific web portal. J. Grid Comput. **13**, 159–175 (2015)
21. C.H. Bennett, G. Brassard, Quantum cryptography: public key distribution and coin tossing, in *Proceedings of IEEE International Conference on Computer Systems and Signal Processing*, New York, vol. 175 (1984), p. 8
22. T. Bjerregaard, S. Mahadevan, A survey of research and practices of network-on-chip. ACM Comput. Surv. **38** (2006). doi:10.1145/1132952.1132953
23. S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, R. Durst, K. Scott, H. Weiss, Delay-tolerant networking: an approach to inter-planetary internet. IEEE Commun. Mag. **41**(6), 128–136 (2003)
24. S. Burleigh, Interplanetary overlay network: an implementation of the DTN bundle protocol, in *Proceedings of 4th IEEE Consumer Communication and Networking Conference* Las Vegas, NV, January 2007, pp. 222–226
25. W.E. Burr, Selecting the advanced encryption standard. IEEE Secur. Priv. **99**, 43–52 (2003)
26. E. Calvalcante, J. Pereira, M. Pitanga et al., On the interplay of internet of things and cloud computing: a systematic mapping study. Comput. Commun. **89–90**, 17–32 (2016)
27. R. Cavallari, F. Martelli, R. Rosini et al., A survey on wireless body area networks: technologies and design challenges. IEEE Commun. Surv. Tutorials **16**(3), 1635–1657 (Third Quarter 2014)
28. P.K. Chandra, A,K. Turuk, B. Sahoo, Survey on optical burst switching in WDM networks, in *Proceedings of the Fourth International Conference on Industrial and Information Systems, ICIIS 2009* (2009), pp. 83–88
29. B.C. Chatterjee, N. Sarma, E.Oki, Routing and spectrum allocation in elastic optical networks: a tutorial. IEEE Commun. Surv. Tutorials **17**(3), 1776–1800 (Third Quarter 2015)
30. S.-J. Chen, Y.-C. Lan, *Reconfigurable Networks-on-Chip* (Springer, New York, NY, 2012)
31. T. Chen, X. Gao, G. Chen, The features, hardware and architectures of data center networks: a survey. J. Parallel Distrib. Comput. **96**, 45–74 (2016)
32. J.M. Chung, H.K. Khan, H.M. Soo, J.S. Reyes, G.Y. Cho, Analysis of GMPLS architectures, topologies and algorithms, in *Proceedings of the 2002 45th Midwestern Symposium on Circuits and Systems MWSCAS-2002*, vol. 3 (2002), pp. 284–287
33. C. Colman-Meixner, C. Develder et al., A survey of resiliency techniques in cloud computing infrastructures and applications. IEEE Commun. Surv. Tutorials **18**(3), 2244–2281 (Third Quarter 2016)
34. J. Daemen, V. Rijmen, The first 10 years of advanced encryption. IEEE Secur. Priv. **6**, 72–74 (2010)
35. J. D'Ambrosia, D. Law, M. Nowell, 40 gigabit ethernet and 100 gigabit ethernet technology overview. Ethernet Alliance (2008). www.ethernetalliance.org
36. J. D'Ambrosia, P. Mooney, 400 Gb/s ethernet: why now? Ethernet Alliance white paper, April (2013)
37. J. D'Ambrosia, 400 gigabit ethernet for the ages. Lightwave, July 10 (2014)
38. D. De Guglielmo, S. Brienza, G. Anastasi, IEEE 802.15.4e: a survey. Comput. Commun. (2016)
39. J.P. Dunning, Taming the blue beast: a survey of bluetooth-based threats. in *IEEE Security and Privacy*, March/April 2010, pp. 20–27
40. C. Eklund, R.B. Marks, K.L. Stanwood, S. Wang, IEEE standard 802.16: a technical overview of the wireless MAN interface for broadband wireless access. IEEE Commun. Mag. **40**, 98–107 (2002)

41. Y. Elkhatib, C. Edwards, Passive network awareness as a means for improved grid scheduling. J. Grid Comput. **13**, 275–291 (2015)
42. K. Fall, A delay-tolerant network architecture for challenged internet, in *Proceedings of SIGCOMM'03*, August 2003, pp. 27–34
43. H. Farhady, H.Y. Lee, A. Nakao, Software-defined networking: a survey. Comput. Netw. **81**, 79–95 (2015)
44. Fierce Wireless, Sprint to shutter WiMax network by end of 2015, will turn off at least 6,000 towers, Fierce Wireless, April 7, 2014
45. J.A. Fraire, P.A. Ferreyra, Assessing DTN architecture reliability for distributed satellite constellations: preliminary results from a case study, in *Proceedings of 2014 Biennial Congress of Argentina (ARGENCON)*, 2014, pp. 564–569
46. J.A. Fraire, J.M. Finochietto, Design challenges in contact plans for disruption-tolerant satellite networks. IEEE Commun. Mag. **53**, 163–169 (2015)
47. T. Garritano, Globus: an infrastructure for resource sharing. Clusterworld **1**(1), 30–31 (2003)
48. R.C. Garroppo, S. Giordano, L. Tavanti, Implementation frameworks for IEEE 802.11s systems. Comput. Commun. **33**, 336–349 (2010)
49. L. Goldberg, 802.16 wireless LANs: a blueprint for the future? Electron. Des. **4**, 44–52 (1997)
50. P. Grun, Introduction to infiniband for end users. Infiniband Trade Association (2010). www.infinibandta.org
51. J.C. Haartsen, S. Mattisson, Bluetooth - a new low-power radio interface providing short-range connectivity. Proc. IEEE **88**(10), 1651–1661 (2000)
52. A. Haque, S.M. Alhashmi, R. Parthiban, A survey of economic models in grid computing. Futur. Gener. Comput. Syst. **27**, 1056–1069 (2011)
53. G.R. Hiertz, D. Denteneer, S. Max et al., IEEE 802.11s: the WLAN mesh standard. IEEE Wirel. Commun. **71**, 104–111 (2010)
54. C. Hoymann, D. Astely et al., LTE release 14 outlook. IEEE Commun. Mag. **54**, 44–49 (2016)
55. F. Hu, Q. Hao, K. Bao, A survey on software-defined network and openflow: from concept to implementation. IEEE Commun. Surv. Tutorials **16**(4), 2181–2206 (Fourth Quarter 2014)
56. T.S. Humble, Quantum security for the physical layer. IEEE Commun. Mag. **51**, 56–62 (2013)
57. A. Hunger, Advanced computer architecture exercise: interconnection networks. presentation (2008)
58. S. Iqbal, L.M. Kiah et al., On cloud security attacks: a taxonomy and intrusion detection and prevention as a service. J. Netw. Comput. Appl. **74**, 98–120 (2016)
59. Y. Jarraya, T. Madi, M. Debbabi, A survey and a layered taxonomy of software-defined networking. IEEE Commun. Surv. Tutorials **16**(4), 1955–1980 (Fourth Quarter 2014)
60. J.M. Kahn, R.H. Katz, K.S.J. Pister, Emerging challenges: mobile networking for "smart dust". J. Commun. Netw. **2**(3), 188–196 (2000)
61. K. Kant, Data center evolution: a tutorial on state on the art, issues and challenges. Comput. Netw. **53**, 2939–2965 (2009)
62. S. Kapp, 802.11: leaving the wire behind. IEEE Internet Comput. **6**(1), 82–85 (2002)
63. C. Kesselman et al., The anatomy of the grid: enabling scalable virtual organizations. J. Supercomput. Appl. **15**(3), 200–225 (2001)
64. M.A. Khan, A survey of security issues in cloud computing. J. Netw. Comput. Appl. **71**, 11–29 (2016)
65. B. Kleyman, Why google wants to rethink data center storage. Data Center Knowledge, May 2 (2016)
66. D. Kreutz, F.M.V. Ramos, P. Verissimo et al., Software-defined networking: a comprehensive survey. Proc. IEEE **103**, 14–76 (2015)
67. G. Ku, J.M. Walsh, Resource allocation and link adaptation in LTE and LTE advanced: a tutorial. IEEE Commun. Surv. Tutorials **17**(3), 1605–1633 (Third Quarter 2015)
68. P. Kumar, S.B. Rana, Development of modified AES algorithm for data security. Optik **127**, 2341–2345 (2016)

69. G. Kumar, E. Jain et al., Mobile cloud computing architecture, application model and challenging issues, in *Proceedings of the 2014 Sixth International Conference on Computational Intelligence and Communication Networks*, 2014, pp. 613–617
70. M.S. Kuran, Y. Tugcu, A survey on emerging broadband wireless access technologies. Comput. Netw. **51**, 3013–3046 (2007)
71. K.S. Kwak, S. Ullah, N. Ullah, An overview of IEEE 802.15.6 standard, in *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, 2010, pp. 1–6
72. R.O. LaMaire, A. Krishnan, P. Bhagwat, J. Panian, Wireless LANs and mobile networking standards and future directions. IEEE Commun. Mag. **34**(8), 86–94 (1996)
73. B. Latré, B. Braem, I. Moerman, C. Blondia, P. Demeester, A survey on wireless body area networks. Wirel. Netw **17**, 1–18 (2011)
74. J. Lawrence, Designing multiprotocol label switching networks. IEEE Commun. Mag. **39**, 134–142 (2001)
75. J. Lee, Y. Kim et al., LTE-advanced in 3GPP Rel-13/14: an evolution towards 5G. IEEE Commun. Mag. **54**, 36–42 (2016)
76. C.E. Leiseron, Fat-tress: universal networks for hardware-efficient supercomputing. IEEE Trans. Comput. **C-34**(10), 892–901 (1985)
77. A. Leon-Garcia, L.G. Mason, Virtual network resource management for next-generation networks. IEEE Commun. Mag. **41**, 102–109 (2003)
78. T. Li, MPLS and the evolving internet architecture. IEEE Commun. Mag. **37**, 38–41 (1999)
79. M. Littlewood, I.D. Gallagher, Evolution toward an ATD multi-service network. Br. Telecom Technol. J. **5**(2), 52–62 (1987)
80. H.J. Mahanta, A. Biswas, M.A. Hussain, Networks on chips: the new trend of on chip interconnection, in *2014 Fourth International Conference on Communication Systems and Network Technologies* (IEEE, New York, 2014)
81. L.O. Mailloux, M.R. Grimaila et al., Performance evaluations of quantum key distribution system architectures. IEEE Secur. Priv. **13**, 30–40 (2015)
82. L.O. Mailloux, C.D. Lewis II et al., Post-quantum cryptography: what adavancements in quantum computing means for IT professionals. IT Pro, Sept/Oct 2016, pp. 42–47
83. A. Majot, R. Yampolskiy, Global catastrophic risk and security implications of quantum computers. Futures **72**, 17–26 (2015)
84. A.G. Malis, MPLS-TP: where are we? OFC/NFOEC Technical Digest, OSA, 2012
85. S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang, A. Ghalsasi, Cloud computing - the business perspective. Decis. Support. Syst. **51**, 176–189 (2011)
86. J.L. Marzo, E. Calle, C. Scoglio, T. Anjali, QoS online routing and MPLS multilevel protection: a survey. IEEE Commun. Mag. **41**, 126–132 (2003)
87. T. Mastelic, A. Oleksiak et al., Cloud computing: survey on energy efficiency. ACM Comput. Surv. **47**(2), 36 pp. (2014). Article 33
88. C.A. Mattmann, J. Garcia, I. Krka et al., Revisiting the anatomy and physiology of the grid. J. Grid Comput. **13**, 19–34 (2015)
89. M. Mauve, H. Hastenstein, A. Widmer, A survey of position-based routing in mobile ad hoc networks. IEEE Netw. **15**(3), 30–39 (2001)
90. C. Meirosu, P. Golonka, A. Hirstius et al., Native 10 gigabit ethernet experiments over long distances. Futur. Gener. Comput. Syst. **21**, 457–468 (2005)
91. C. Messerschmidt, O. Hinz, Explaining the adoption of grid computing: an integrated institutional theory and organizational capability approach. J. Strateg. Inf. Syst. **22**, 137–156 (2013)
92. R.M. Metcalfe, D.R. Boggs, Ethernet: distributed packet switching for local computer networks. Commun. ACM **19**, 395–404 (1976)
93. C. Metz, C. Barth, C. Filsfils, Beyond MPLS. . . less is more. IEEE Internet Comput. **11**, 72–76 (2007)
94. T.P. Morgan, A new age in cluster interconnects dawns. The Next Platform, November 22 (2015)

95. J. Moy, OSPF version 2 RFC 2178, April (1998)
96. J. Mukherjee, B. Ramamurthy, Communication technologies and architectures for space network and interplanetary internet. IEEE Commun. Surv. Tutorials **15**(2), 881–897 (Second Quarter 2013)
97. W.J. Munro, K. Azuma, Inside quantum repeaters. IEEE J. Sel. Top. Quantum Electron. **21**(3), (2015)
98. S.L. Muringi, Behind the scenes with dense wavelength division multiplexing (DWDM). TechnoMag, May 18 (2016)
99. C.S.R. Murthy, B.S. Manoj, *Ad Hoc Wireless Networks: Architectures and Protocols* (Prentice-Hall, Upper Saddle River NJ, 2004)
100. New release of infiniband trade association architecture specification, press release, edited by StorageNewsletter.com, March 19 (2015)
101. M. Nowell, V. Vusirikala, R. Hays, Overview of requirements and applications for 40 gigabit and 100 gigabit ethernet. Ethernet Alliance, version 1.0, www.ethernetalliance.org (2007)
102. U.Y. Ogras, R. Marculescu, *Modeling, Analysis and Optimization of Network-on-Chip Communication Architectures* (Springer, Dordrecht, 2013)
103. Old standard new topic - optical transport network, FS.Com, July 14, 2016
104. R.O. Onvural, *Asynchronous Transfer Mode Networks: Performance Issues*, 2nd edn. (Artech House, Norwood, 1995)
105. R.M. Pacella, Hacking the cloud Popular Science, April 2011, pp. 68–72
106. M. Palesi, M. Daneshtalab (eds.), *Routing Algorithms in Networks-on-Chip* (Springer, New York, NY, 2014)
107. N. Panwar, S. Sharma, A.K. Singh, A survey on 5G: the next generation of mobile communication. Phys. Commun. **18**, 64–84 (2016)
108. D. Papadimitriou, D. Verchere, GMPLS user-network interface in support of end-to-end rerouting. IEEE Commun. Mag. **43**, 35–43 (2005)
109. N. Papanikolaou, An introduction to quantum cryptography. Crossoroads **11**(3), 3 (ACM, New York, 2005)
110. I. Papapanagiotou, D. Toumpakaris, J. Lee, M. Devetsikiotis, A survey on next generation mobile WiMAX networks: objectives, features and technical challenges. IEEE Commun. Surv. Tutorials **11**(4), 3–18 (2009)
111. S. Parkes, P. Armbruster, SpaceWire: spacecraft onboard data-handling network. Acta Astronautica **66**, 88–95 (2010)
112. S. Parkes, C. McClements, D. McLaren et al., SpaceFibre: adative high-speed data-link for future spacecraft onboard data handling, in *2014 NASA/ESA Conference on Adaptive Hardware and Systems (AHS)*, 2014, pp. 164–171
113. S. Parkes, C. McClements, D. McLaren et al., SpaceFibre: A Multi-Gigabit/s interconnect for spacecraft onboard data handling, in *Proceedings of 2015 IEEE Aerospace Conference*, 2015, pp. 1–13.
114. C.E. Perkins, *Ad Hoc Networks* (Addison-Wesley, Boston, MA, 2001)
115. S.W. Peters, R.W. Heath Jr., The future of WiMAX: multihop relaying with IEEE 802.16j. IEEE Commun. Mag. 104–111 (2009)
116. I. Poole, What exactly is 802.11n? Commun. Eng. 46–47 (2007)
117. I. Poole, What exactly is. . . LTE? Commun. Eng. 46–47 (2007)
118. J.M. Rabaey, M.J. Ammer, J.L. da Silva Jr. et al., PicoRadio supports ad hoc ultra-low power wireless networking. Computer **33**(7), 42–48 (2000)
119. M. Ricknas, As LTE-advanced becomes more commom, 4G speeds increase. Computerworld, January 9, 2015
120. M. Rinne, O. Tirkkonen, LTE, the radio technology path towards 4G. Comput. Commun. **33**, 1894–1906 (2010)
121. T.G. Robertazzi, Peformance evaluation of high speed switching fabrics and networks: ATM, broadband ISDN and MAN technology. IEEE Press, 1993 (now distributed by Wiley)
122. T.G. Robertazzi, *Networks and Grids: Technology and Theory* (Springer, New York NY, 2007)
123. T.G. Robertazzi, *Basics of Computer Networking* (Springer, New York, 2011)

124. H. Rong, H. Zhang et al., Optimizing energy consumption for data centers. Renew. Sust. Energ. Rev. **58**, 674–691 (2016)
125. J.P. Ryan, WDM: North American deployment trends. IEEE Commun. Mag. **36**(2), 40–44 (1998)
126. R.K. Saini, M. Ahmed, 2D hexagonal mesh vs 3D mesh network on chip: a performance evaluation. Int. J. Comput. Digit. Syst. **4**(1), 33–41 (2015)
127. T. Sandholm, D. Lee, Notes on cloud computing principles. J. Cloud Comput. **3**, 21 (2014)
128. V. Scarani, C. Kurtsiefer, The black paper of quantum cryptography: real implementation problems. Theor. Comput. Sci. **560**, 27–32 (2014)
129. D. Schneider, Q. Hardy, Under the hood at google and facebook. IEEE Spectr. **48**(6), 63–67 (2011)
130. J.M. Schopf, B. Nitzberg, Grids: the top ten questions Sci. Program. **10**(2), 103–111 (IOS Press, New York, 2002)
131. W. Schuchart, Watt's up? DatacenterDynamics, July 17, 2015
132. M. Schwartz, *Telecommunication Networks: Protocols, Modeling and Analysis* (Addison-Wesley, Reading, MA, 1987)
133. L. Schwiebert, S.K.S. Gupta, J. Weinmann, Research challenges in wireless networks of biomedical sensors, in *ACM Sigmobile*, pp. 151–165, 2001
134. U. Schwiegelshohn, R.M. Badia et al., Perspectives on grid computing. Futur. Gener. Comput. Syst. **26**, 1104–1115 (2010)
135. R.Q. Shaddad, A survey on access technologies for broadband optical and wireless networks. J. Netw. Comput. Appl. **41**, 459–472 (2014)
136. R.C. Shah, J.M. Rabaey, Energy aware routing for low energy ad hoc networks, in *Proceedings of the 3rd IEEE Wireless, Communications and Networking Conference*, pp. 350–355, 2002
137. N. Shaji, P.L. Bonifus, Design of AES architecture with area and speed tradeoff. Procedia Technol. **24**, 1135–1140 (2016)
138. Y. Sharma, B. Javadi et al., Reliability and energy efficiency in cloud computing systems: survey and taxonomy. J. Netw. Comput. Appl. **74**, 66–85 (2016)
139. P.W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J. Comput. **26**(5), 1484–1509 (1997)
140. C.E. Siller (ed.), *SONET/SDH: A Sourcebook of Synchronous Networking* (Wiley-IEEE Press, New York, 1996)
141. M. Singh, M.L. Singh, A novel algorithm to integrate synchronous digital hierarchy into optical transport network using mixed line rates. Optiks **125**, 6739–6745 (2014)
142. S. Singh, Y.-S. Jeong, J.H. Park, A survey of cloud security: issues, threats and solutions. J. Netw. Comput. Appl. **75**, 200–222 (2016)
143. S. Siwamogsatham, 10 Gbps ethernet (1999). www.cse.wustl.edu/jain
144. W. Stallings, *High-Speed Networks and Internets: Performance and Quality of Service* (Prentice-Hall, Upper Saddle River, 2002)
145. N. Sultan, Cloud computing for education: a new dawn. Int. J. Inf. Manag. **30**, 109–116 (2010)
146. W. Sun, O. Lee, Y. Shin et al., Wi-Fi could be much more. IEEE Commun. Mag. **24**, 22–29 (2014)
147. A. Suwan, F. Siewe, N. Abwnawar, Towards monitoring security policies in grid computing, in *SAI Computing Conference* (2016), pp. 573–578
148. A. Tanenbaum, *Computer Networks* 3rd edn. (Prentice-Hall, Upper Saddle River, 1996)
149. A. Tanenbaum, *Computer Networks*, 4th edn. (Prentice-Hall, Upper Saddle River, 2003)
150. K. Tatas, K. Siozios et al., *Designing 2D and 3D Network-on-Chip Architectures* (Springer, New York, NY, 2014)
151. M. Tatipamula, F. Le Faucheur, T. Otani, H. Esaki, Implementation of IPv6 services over a GMPLS-based IP/optical network. IEEE Commun. Mag. **43**, 114–122 (2005)
152. I. Tomkos, S. Azodolmolky et al., A tutorial on the flexible optical networking paradigm: state of the art, trends and research challenges. Proc. IEEE **102**(9), 1317–1337 (2014)
153. R. Triggs, What is LTE advanced? Android Authority, March 2 (2016)

154. S.J. Vaughan-Nichols, Will 10-gigabit ethernet have a bright future? Computer **35**, 22–24 (2002)
155. S.J. Vaughan-Nichols, Will the new wi-fi fly? Computer **39**, 16–18 (2006)
156. E. van der Linde, G.P. Hancke, An investigation of bluetooth mergence with ultra wideband. Ad Hoc Netw. **9**(5), 852–863 (2011)
157. J. Wannstrom, LTE-advanced, 3GPP.org, June 2013
158. J.S. Ward, A. Baker, Observing the clouds: a survey and taxonomy of cloud monitoring. J. Cloud Comput. **3** (2014). doi:10.1186/s13677-014-0024-2
159. What is optical transport network? MetaSwitch Networks (undated)
160. S. Wiesner, Conjugate coding. Sigarc News **15**(1), 78–88 (1983) (original manuscript written about 1969)
161. A.E. Willner, Communication with a twist. IEEE Spectr. **35**, 35–39 (2016)
162. J. Williams, The 802.11b Security Problem - Part I, IEEE ITPro, pp. 91–96 (2001)
163. R. Winter, The coming of age of MPLS. IEEE Commun. Mag. **49**, 78–81 (2011)
164. A. Woodie, Does infiniband have a future on hadoop? Datanami, August 4 (2015)
165. W. Xia, Y. Wen, C.H. Foh et al., A survey of software-defined networking. IEEE Commun. Surv. Tutorials **17**(1), 27–51 (First Quarter 2015)
166. X. Xiao, L.M. Ni, Internet QoS: a big picture. IEEE Netw. **13**, 8–18 (1999)
167. K. Yamamoto, New infiniband architecture specification offers improved scalability and management, Tom's IT PRO, March 10 (2015)
168. D. Yang, Y. Xu, M. Gidlund, Coexistence of IEEE802.15.4 based networks: a survey, in *Proceedings of the 36th Annual IEEE Conference on Industrial Electronics (IECON 2010)*, 2010. pp. 2107–2113
169. Q. Yu, X. Sun, R. Wang et al., The effect of DTN custody transfer in deep-space communications. IEEE Wirel. Commun. **169**, 169–176 (2013)
170. J. Zhang, H. Huang, X. Wang, Resource provision algorithms in cloud computing: a survey. J. Netw. Comput. Appl. **64**, 23–42 (2016)
171. J. Zheng, M.J. Lee, Will 802.15.4 make ubiquitous networking a reality? a discussion on a potential low power, low bit rate standard. IEEE Commun. Mag. **42**, 140–146 (2004)
172. D. Zissis, D. Lekkas, Addressing Cloud Computing Security Issues. Futur. Gener. Comput. Syst. **28**(3), 583–592 (2011)

Index

T.G. Robertazzi, *Introduction to Computer Networking*,
DOI 10.1007/978-3-319-53103-8